

今回の講座の目的は『世の中にあふれた数字に価値を見出す手法を学ぶ』です。

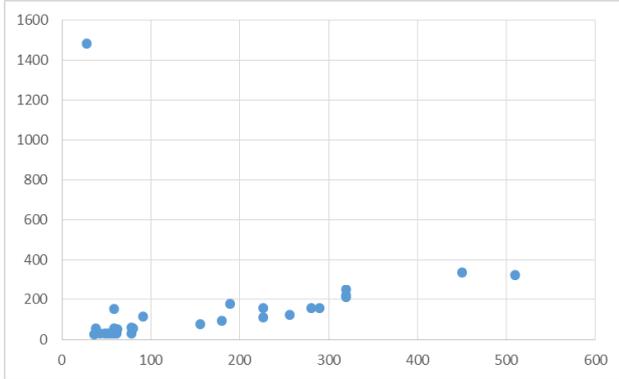
本時は、回帰直線を学びます。

**【回帰直線】**  
 回帰直線とは、散布図において、予定値を求める際に用いられる直線のこと。2組のデータの中心的な分布傾向を表すもので、最小二乗法と呼ばれる算術を用いて求められる。

それでは、Excelを用いて、dummydateAの大人の来場者数と子どもの来場者数の回帰直線を描いてみましょう。

- 手順①：前回学んだ手順で、相関図を作成する。
- 手順②：グラフツールにある『グラフ要素追加』をクリックする。
- 手順③：『近似曲線』にカーソルをあわせ、『線形』をクリックする。  
 ※線の色を変えたければ、線の上にカーソルを合わせ、右クリックして枠線の色を変えると変更できる。

(1) 手順③で得られたグラフを下に書き写せ。



次に、得られた回帰直線に数式を表示してみましょう。

- 手順④：グラフの近似曲線を左クリックし、右クリックした後、近似曲線の書式設定をクリックする。
- 手順⑤：近似曲線のオプション下方にある『グラフに数式を表示する』をチェックする。

(2) 手順⑤で得られた数式を答えなさい。

ここで、データサイエンスしましょう！  
 dummydateAの大人と子供の来場者数には、統計上問題点があったと思います。それを踏まえて、正しい回帰直線の式を求めてみましょう。

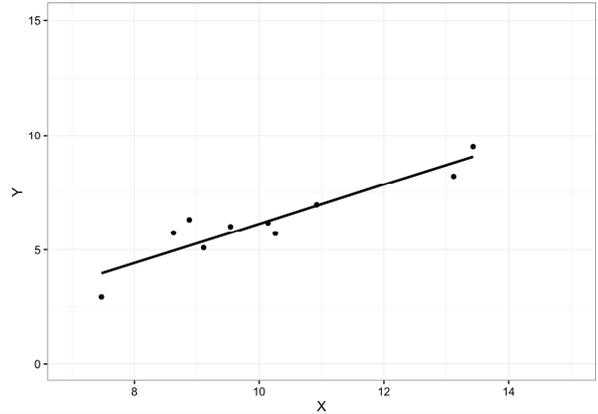
**【正しい回帰直線の式】**

この正しい回帰直線の式を用いて、次の問に答えましょう。

(3) 大人が400人来場したとすると、子どもは何人来場することが見込まれるかを答えよ。

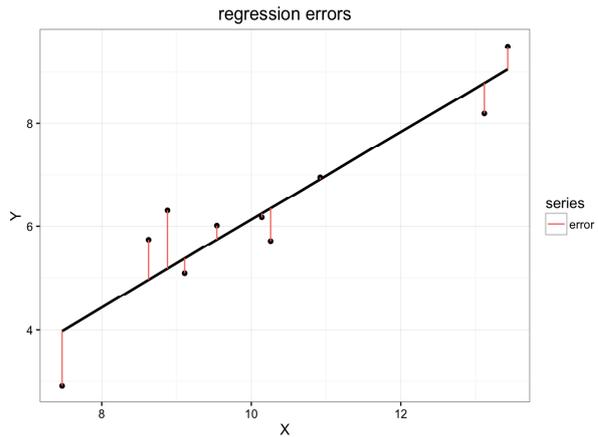
最後に、回帰直線がどのようにしてできているのかを学びます。

次のような散布図及び回帰直線があったとします。



※X軸の値を予測する変数とし、原因変数と呼ばれます。  
 ※Y軸の値を予測される変数とし、結果変数と呼ばれます。

回帰直線とは、YとXの関係を一次関数  $y = ax + b$  によって説明するための式です。y切片  $b$  は結果変数が最小の場合のYの値としたいですし、傾き  $a$  を説明が適当にできるように定めたいです。そこで、最小二乗法という考え方を活用します。



上図のように、直線と各データのy軸方向の距離を考えます。それを2乗し、全て加えたものが最小となれば、最も説明に適した直線になるといえます。この考え方を最小二乗法といい、得られた値を最小二乗基準といいます。  
 この最小二乗基準を、y切片と傾き(回帰係数)について偏微分し、導出される方程式を整理していくことで、最終的に以下のように簡単に算出することができます。

傾き  $a$  (回帰係数) =  $\frac{XとYの共分散}{Xの分散}$

y切片  $b$  = (Yの平均) - (回帰係数) × (Xの平均)

※この回帰係数の式から考えられることを、グループで話し合ってください。