

年 組 番 氏名

SS 理数探究 I 統計処理④(散布図と相関関係)

○データの散布図から関係を推測する

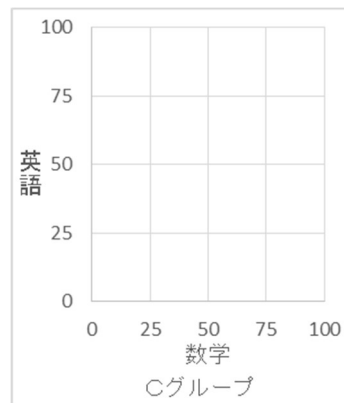
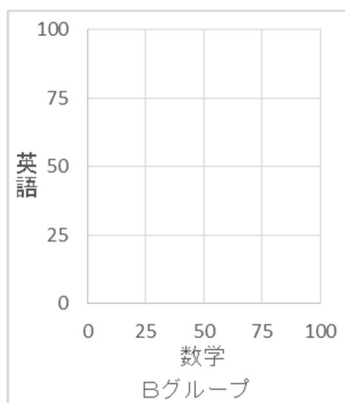
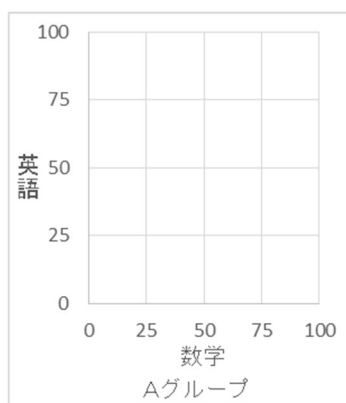
5人ずつA、B、Cの3グループで、数学と英語の試験をして次の表の点数を得ました。

Aグループ	数学	英語
a	75	100
b	50	75
c	50	75
d	50	50
e	25	25
平均	50	65
標準偏差	14.4	23.3

Bグループ	数学	英語
f	50	75
g	50	50
h	50	25
i	50	0
j	50	0
平均	50	30
標準偏差	0.0	26.6

Cグループ	数学	英語
k	100	25
l	100	50
m	50	50
n	0	75
o	0	100
平均	50	60
標準偏差	40.8	23.3

ここで数学の点数を x 、英語を y として、点 (x, y) を次の座標平面上にかいてみよう。



このような図を**散布図**といいます。

問1 この散布図より3つのグループについて、それぞれどのようなことが分かりますか。

解答：

データは1種類の項目で調べても価値はありますが、2種類に増やすことで新しい分析結果を得ることができます。

データの分布がAグループのように右上がりになっているとき**正の相関関係がある**といい、Cグループのように右下がりになっているとき**負の相関関係がある**といいます。このどちらでもないような場合は、**相関関係はない**といいます。

相関の強さを表す相関係数 r というのがあり正の相関の場合は直線に近いほど r は1の値に近づきます。エクセルで相関係数を求めるには CORREL(配列 1、配列 2)を使います。

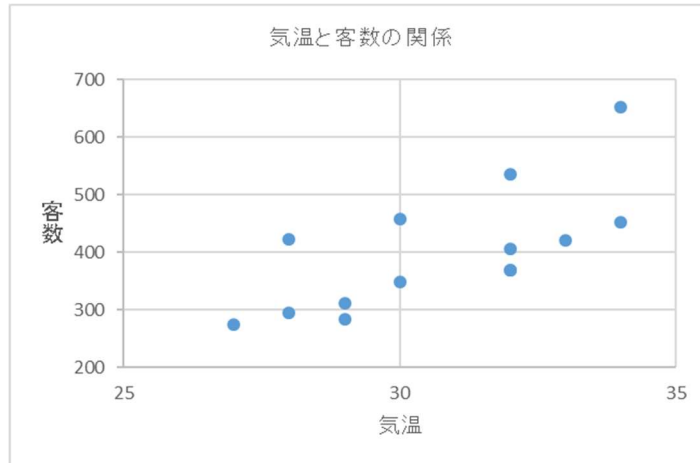
相関係数の求め方の数式は数学 I で詳しく学ぶのでここでは省略します。

統計処理④

○散布図からデータの関係を予測しよう。

例 1 次のデータは、あるアイスクリーム店の 8 月 1 日から 8 月 14 日までの最高気温と客数のデータとその散布図です。

日付	曜日	最高気温	客数
1	月	29	312
2	火	30	348
3	水	29	284
4	木	32	369
5	金	33	420
6	土	32	536
7	日	34	652
8	月	27	275
9	火	28	294
10	水	32	368
11	木	34	451
12	金	32	405
13	土	30	458
14	日	28	422



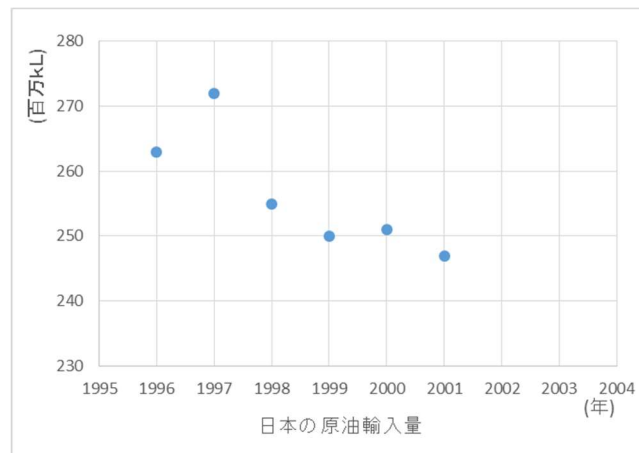
問 2 グラフより全体的にどのようなことがわかりますか。また、上下 2 つのデータに分かれているように見えるのはなぜだと考えられますか。

解答：

○データの散布図から未来を予想する

例 2 日本は大量の原油を輸入することによって国民の生活を維持しています。資料(「読売年鑑 2003」, 読売新聞社)によると、日本の 1996~2001 年における 1 年ごとの原油輸入量は、次の表のように推移したとのことでした。散布図は次のようになります。

日本の原油の輸入量	
年	輸入量(百万kL)
1996	263
1997	272
1998	255
1999	250
2000	251
2001	247



問 3 2004 年の原油の輸入量はどれくらいになると予測されるでしょうか。また、どのようにすればそれが予測できますか。

解答：

統計処理④

※以下の問は、時間があればエクセルでやってみよう。

問4 例1の散布図をエクセルでかいてみよう。

また、相関係数をエクセルで求めなさい。相関係数の式はCORREL(配列1、配列2)

解答：エクセルで散布図をかくことができたなら○を記入しよう。→ ()

相関係数は ()

- ・ 散布図を作成する。(まず、左上の表を作成してデータを入力する)
 - ①作成した表の「最高気温」と「客数」のデータを、見出しを含めて選択します。
 - ②「挿入」タブの「グラフ」→「散布図」を選択
- ・ タイトルや軸の見出しを指定する。
 - ①散布図をクリックして、右の+の「軸ラベル」をチェック
 - ②グラフの縦軸と横軸に「軸ラベル」と表示されるので、横軸のラベルに「気温」と入力。
 - ③縦軸の「軸ラベル」を右クリックし「軸ラベルの書式設定」→右の「サイズとプロパティ」→「文字列の方向」→「縦書き」。その後、縦軸のラベルに「客数」と入力。
 - ④タイトルをクリックして選択し、「気温と客数の関係」と入力。
- ・ 軸の最小値・最大値・目盛間隔を指定する。
 - ①散布図の横軸の数字を選択し右クリック→「軸の書式設定」→最小値「25」、最大値「35」、目盛「5」と入力する。
 - ②同様に、縦軸も最大値を「700」、最小値を「200」、目盛間隔を「100」と指定する。

実は、例2において6つの点のなるべく近くを通る直線を計算で求めることができます。

その直線を**回帰直線**と言います。回帰直線はエクセルを用いて簡単に求められます。

- ・ エクセルによる回帰直線の求め方
 - ①エクセルに表を作成し、その散布図を作る。
 - ②作成した散布図のデータ点をクリックして、右の+の「近似曲線」→「その他のオプション...」の順に進む。
 - ③「近似曲線の書式設定」画面が現れるので、「線形近似(L)」のグラフを選択し、「グラフに数式を表示する(E)」の前にチェックを入れる。
 - ④グラフの直線の式を紙に書き写して、求めたいxの値を代入してyの値を求める。

問5 例2の散布図をエクセルでかき、回帰直線を求めて、2004年の輸入量を予測せよ。

解答：

統計処理④

補足：外れ値に注意！

外れ値とは、全体の分布の中心から極端に外れた値のことです。散布図を観察することによって外れ値を発見することができます。データの入力間違いの場合は修正すればよいのですが、そうでない場合は、外れ値の扱いを慎重に検討しなくてはなりません。外れ値をデータ分析から除外するかどうかは、研究目的と照らしあわせて検討することになります。