

データの分析

「情報II」第7章

Contents

1. クロス集計	01
2. アンケート集計	05
3. 回帰分析	09
4. 外部参照関数	15
5. アンケート分析	17

●本書の複製等について—本書のコピー、スキャン、デジタル化等の無断複製は著作権法上での例外を除き禁じられています。本書を代行業者等の第三者に依頼してスキャンやデジタル化することは、たとえ個人や家庭内の利用でも認められておりません。

クラス： 番号： 氏名：

クロス集計

データベースのSQLによる関係演算では、グループごとの集計等はできましたが、複数の項目にまたがった集計は簡単にはできません。そこで、表計算アプリを使ったクロス集計を行うことで、複数項目にまたがった集計を簡単に行うことができます。

(教科書I : p.204 – p.205)

■ クロス集計とは

クロス集計とは

クロス集計 = 2つ以上のカテゴリ間でデータを比較するための集計方法

たとえば、アンケートで興味があるかどうかと、年齢層をそれぞれ聞いたとする

→単純集計では、年齢層、興味の有無のそれぞれ合計しかわからない

→クロス集計をすると、年齢層と興味の有無をかけあわせて集計することができる

	興味がある	どちらでもない	興味がない	総計
中高生	57	24	14	95
若者	44	26	32	102
中年	32	32	24	88
高齢者	14	24	31	69
総計	133	106	101	340

クロス集計のしくみ

左の表から、性別と文理別をかけあわせて集計すると、右の表のようになる

名前	性別	文理別
太郎	男性	文系
花子	女性	理系
聡美	女性	理系
健太	男性	文系
智子	女性	文系
亮介	男性	理系

性別	文系	理系	総計
女性	1	2	3
男性	2	1	3
総計	3	3	6

※2つの項目の値を見出しにし、それぞれの交わるところに件数や合計値などを入れる


■ クロス集計の方法

ピボットテーブルの追加

- ※表計算アプリ上では「**ピボットテーブル**」と呼ばれる
- ※説明として右の表を例に説明していく

名前	性別	文理別
太郎	男性	文系
花子	女性	理系
聡美	女性	理系
健太	男性	文系
智子	女性	文系
亮介	男性	理系

ピボットテーブルの作成

- ①表選択中にメニューの「挿入→ピボットテーブル」
- ②挿入先を「新しいシート」にして[作成]を押す

ピボットテーブルエディタの使い方

- ①列見出しを、「行」「列」「値」にドラッグ
- ②「値」の集計方法を変更する

名前 ×

集計 表示方法

COUNTA デフォ...

※主な集計方法は以下の通り

合計	SUM
平均	AVERAGE
カウント	COUNTA
中央値	MEDIAN
最大値	MAX
最小値	MIN
標準偏差	STDEVP

ピボットテーブル エディタ ×

シート1!A1:C13 検索

候補

行

列

値

フィルタ

名前

性別

文理別

追加

追加

追加

追加

列見出しをドラッグ

行 追加

性別 ×

並べ替え 並べ替え

昇順 性別

総計を表示

列 追加

文理別 ×

並べ替え 並べ替え

昇順 文理別

総計を表示

値 追加

名前 ×

集計 表示方法

COUNTA デフォ...

クロス集計の方法

右のように設定をすると、次のように対応する

名前	性別	文理別	性別	文系	理系	総計
太郎	男性	文系				
花子	女性	理系				
聡美	女性	理系				
健太	男性	文系				
智子	女性	文系				
亮介	男性	理系				
			女性	1	2	3
			男性	2	1	3
			総計	3	3	6

ワークシート

問題1

次のデータは、野球選手の所属チーム、利き手、ホームラン数を一覧にしたものです。

連番	選手	チーム	利き手	HR
1	ひらの	チームD	左	10
2	かみかわ	チームD	左	6
3	うの	チームD	右	32
4	げいり	チームD	右	34
5	おおしま	チームD	右	24
6	かわまた	チームD	左	10
7	なかお	チームD	右	12
8	やすとも	チームD	右	10
9	まつもと	チームG	左	4
10	しのすか	チームG	左	6
11	くろまで	チームG	左	36
12	たつのり	チームG	右	32
13	よしむら	チームG	左	22
14	なかはた	チームG	右	16
15	こうの	チームG	右	4
16	やまくら	チームG	右	10
17	いしげ	チームL	右	28
18	かなもり	チームL	左	8
19	あきやま	チームL	右	40
20	きよはら	チームL	右	32
...

①ホームラン数のもっとも多いチームはどこですか？

②左利きの選手がもっとも多いチームはどこですか？

③右利きの選手と左利きの選手では、どちらの方が平均ホームラン数が多いですか？

④平均ホームラン数をもっとも多いのは、どのチームの利き手がどちらの選手ですか？

問題2

次のデータは、ある百貨店における商品の販売データの一部です。次のそれぞれの項目について、もっとも多く記録されている組み合わせを教えてください。

顧客ID	性別	年齢層	購入商品	HR
1	男性	50-59歳	家電	キャッシュレス決済
2	男性	20-29歳	家具	クレジットカード
3	女性	30-39歳	家電	現金
4	男性	50-59歳	食品	クレジットカード
5	女性	30-39歳	家電	キャッシュレス決済
6	男性	40-49歳	食品	キャッシュレス決済
7	男性	40-49歳	家具	現金
8	女性	50-59歳	衣料品	キャッシュレス決済
9	男性	50-59歳	食品	キャッシュレス決済
10	女性	20-29歳	家具	現金
11	女性	30-39歳	家電	クレジットカード
12	男性	20-29歳	衣料品	キャッシュレス決済
13	女性	20-29歳	衣料品	クレジットカード
14	男性	30-39歳	食品	現金
15	女性	50-59歳	食品	クレジットカード
16	女性	30-39歳	家具	現金
17	女性	30-39歳	家具	クレジットカード
18	女性	50-59歳	衣料品	クレジットカード
19	女性	50-59歳	家具	現金
20	女性	50-59歳	衣料品	キャッシュレス決済
...

①年齢層と購入商品カテゴリー

②年齢層と支払い方法

③購入商品カテゴリーと支払い方法

④性別と購入商品カテゴリー

⑤性別と支払い方法

問題3

次のデータは、ある期間の各店舗における販売製品の数量と売上金額を表しています。

No.	日付	店舗	製品名	単価	数量	売上金額
1	04/01	東口店	ハーブティー	3150	150	472500
2	04/01	東口店	チョコレート	4200	100	420000
3	04/01	東口店	クッキー	3150	100	315000
4	04/01	東口店	有機ジャム	5250	80	420000
5	04/01	西口店	ハーブティー	3150	150	472500
6	04/10	西口店	有機コーヒー	5250	100	525000
7	04/10	西口店	クッキー	3150	150	472500
8	04/10	西口店	ハーブティー	3150	50	157500
9	04/10	南口店	有機コーヒー	5250	82	430500
10	04/15	南口店	有機紅茶	4200	200	840000
11	04/15	南口店	クッキー	3150	100	315000
12	04/15	南口店	ハーブティー	3150	220	693000
13	04/16	西口店	有機ジャム	5250	130	682500
14	04/16	南口店	チョコレート	4200	200	840000
15	04/16	南口店	有機ジャム	5250	100	525000
16	04/16	西口店	有機紅茶	4200	130	546000
17	04/28	西口店	チョコレート	4200	120	504000
18	04/28	西口店	クッキー	3150	150	472500
19	05/01	西口店	有機紅茶	4200	80	336000
20	05/01	西口店	有機紅茶	4200	50	210000
...

- ① 売上の合計金額を日付と店舗ごとに集計する表を作成し、5/13の東口店の売上金額を求めてください。
- ② 日付と製品名ごとに製品の販売数量の合計を集計する表を作成し、5/21に売れたチョコレートの販売個数を求めてください。
- ③ 店舗と製品名ごとに販売数量の合計を集計する表を作成し、もっとも多くの数が売れた店舗と製品の組み合わせは何で、それはいくつであったかを求めてください。

振り返り

次の各観点が達成されていれば□を塗りつぶしましょう。

クロス集計とはどのようなものかを理解することができた

様々な観点でクロス集計を行なうことができた

今日の授業を受けて思ったこと、感じたこと、新たに学んだことなどを書いてください。

.....

.....

.....

アンケート集計

自習時間にアンケートの作成と回答を行っていただきました。各自が作成したアンケート結果を集計を行ってみましょう。集計の際、思うように入力されていないデータを正しく整理しなければならないかもしれません。

(教科書I : p.202 – p.203 , 教科書II : p.58 – p.61)

■ 尺度

尺度

<p>→ 3</p> <p style="text-align: center;">都道府県番号</p> <p>26 京都府</p> <p>28 兵庫県</p> <p>27 大阪府</p> <p>30 和歌山県</p> <p>25 滋賀県</p> <p>29 奈良県</p> <p>24 三重県</p> <p>番号自体には意味はない</p>	<p>4</p> <p style="writing-mode: vertical-rl; text-orientation: upright;">英検各級の目安</p> <table border="1"> <tbody> <tr> <td>1級</td> <td>大学上級 程度</td> </tr> <tr> <td>準1級</td> <td>大学中級 程度</td> </tr> <tr> <td>2級</td> <td>高校卒業 程度</td> </tr> <tr> <td>準2級</td> <td>高校中級 程度</td> </tr> <tr> <td>3級</td> <td>中学卒業 程度</td> </tr> <tr> <td>4級</td> <td>中学中級 程度</td> </tr> <tr> <td>5級</td> <td>中学初級 程度</td> </tr> </tbody> </table>	1級	大学上級 程度	準1級	大学中級 程度	2級	高校卒業 程度	準2級	高校中級 程度	3級	中学卒業 程度	4級	中学中級 程度	5級	中学初級 程度
1級	大学上級 程度														
準1級	大学中級 程度														
2級	高校卒業 程度														
準2級	高校中級 程度														
3級	中学卒業 程度														
4級	中学中級 程度														
5級	中学初級 程度														
<p>識別のために数字を付与したものの データの例：背番号、郵便番号 等</p>	<p>相対的な順序に意味のある量 データの例：順位、等級、評定値 等</p>														
<p>↖ 5</p> <p style="text-align: center;">平均気温の変化</p> <p>30</p> <p>25</p> <p>20</p> <p>15</p> <p>6月は5月に比べて 平均気温が5℃上昇</p> <p>基点が0でなくてもよい</p> <p>10</p> <p>5月 6月 7月 8月 9月 10月</p>	<p>6</p> <p style="text-align: center;">果物の重さ比較</p> <p>300</p> <p>250</p> <p>200</p> <p>150</p> <p>100</p> <p>50</p> <p>0</p> <p>オレンジの重さは キウイの2倍</p> <p>基点が0である</p> <p>キウイ オレンジ カキ リンゴ</p>														
<p>たがいの数値の差に意味のある量 データの例：気温、指数、日付 等</p>	<p>たがいの数値の比に意味のある量 データの例：長さ、面積、金額 等</p>														

■ データの整理

データクレンジング

データクレンジング = データの表記のゆれや誤記、重複などを修正し一貫性を向上

※アンケートでは、必ずしも設問の指示通りに答えてくれるとは限らない

例

「平日に授業以外で何分勉強しますか？（半角数字）」という質問があったとする

→半角数字で数値が入力されていることで計算が可能になる

→次のような表記があると、うまく計算できない → 修正が必要

1 2 0	全角文字で書かれた数字は計算ができない
120分	後ろに単位を書きってしまったことで、文字列扱いになっている
120くらい	単位同様、文字が書いてあると文字列扱いになる
1 2 0分	数字も全角、単位も書いてある
2	分で答えて欲しいのに2時間と思われる回答 → 真意不明のため修正不能

外れ値

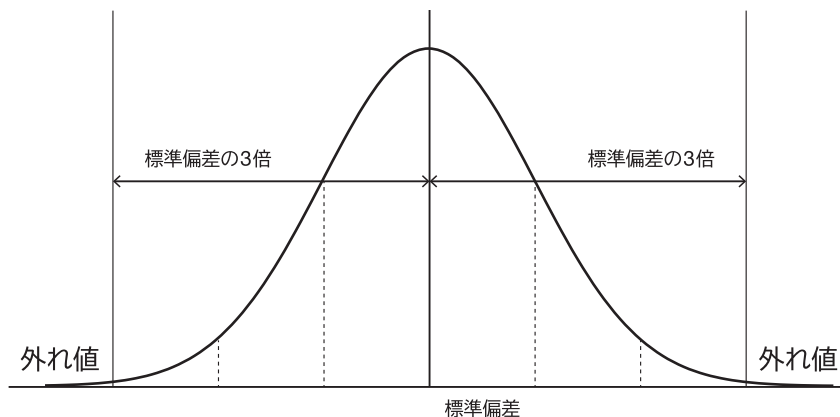
外れ値 = 明らかにほかとはかけ離れた値

※外れ値を処理する方法のひとつは、そのまま取り除くことが考えられる

外れ値の判定

平均から**標準偏差のn倍**以上外れた値を外れ値としてみなす方法がよく用いられる

nは、2倍や3倍が使われることが多い

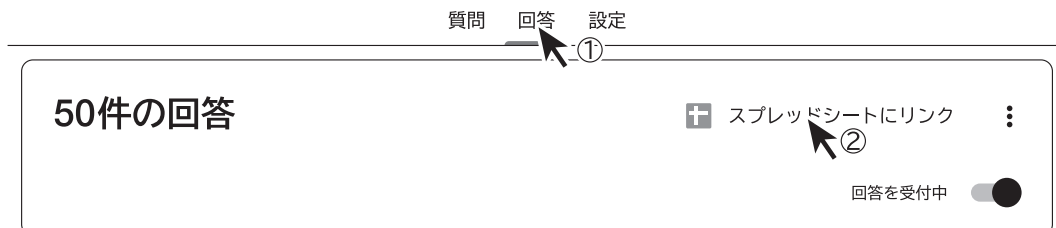


■ アンケート集計

アンケート集計

Google フォームでとった回答を Spreadsheet で集計できるようにする

→①フォームの [回答] タブを開き、②「スプレッドシートにリンク」をクリック



データクレンジング

回答のコピー

アンケートの回答のスプレッドシートの右端の列の右隣に回答をコピーアンドペースト

→①列番号をクリックして選択し、②キーボードの「ctrl+C」 ⇒ **コピー**

→③列番号をクリックして選択し、④キーボードの「ctrl+V」 ⇒ **ペースト**

	A	B	C	D	E	F
	Form_Responses1					
1	タイムスタンプ	メールアドレス	平日の勉強時間 (分)	平日のゲーム時間 (分)		
2			② ctrl + C		④ ctrl + V	
3						

重複行の削除

同じメールアドレスから複数の回答があった場合、片方の行を削除しよう

データの整形

集計をしやすいように、下記のような操作を行おう（コピーしたデータで行うこと）

- ◆ 全角数字を半角数字に
- ◆ 単位等を消して数値のみに
- ◆ 外れ値を削除する
- ◆ 文字データは集計しやすいように統一する

クロス集計

アンケート結果をさまざまな観点でクロス集計する

→件数を数える、平均値・中央値・標準偏差などの統計値を求める 等

課題1

アンケート結果を集計して、どのようなことがわかったかをレポートにして書こう

アンケート結果の共有

アンケート結果を共有するために、スプレッドシート画面右上の「共有」をクリック

→ [一般的なアクセス] を「近江兄弟社高等学校」に設定 ※権限は【閲覧者】でよい

→ 「リンクをコピー」をクリックすると、共有用のURLをコピーできる



課題2

アンケート結果のURLと、数値データの入っている列をフォームに入力してください。

振り返り

次の各観点が達成されていれば□を塗りつぶしましょう。

- データの尺度について、意味を理解することができた
- アンケート結果を集計するためにデータクレンジングが必要なことがわかった
- アンケート結果のデータクレンジングを行うことができた
- アンケート結果を集計し、結果の分析を行うことができた

今日の授業を受けて思ったこと、感じたこと、新たに学んだことなどを書いてください。

.....

.....

回帰分析

たくさんのデータが得られると、そこにどのような傾向があるかを読み取ることができます。ここでは、複数の項目にまたがる量的なデータからどのような傾向が読み取ることができるか、データの分析方法について学習します。

(教科書I : p.194 – p.195 , p.204 – p.205 , 教科書II : p.78 – p.79)

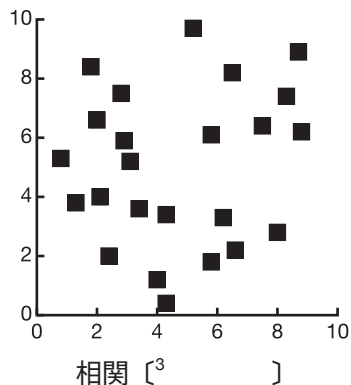
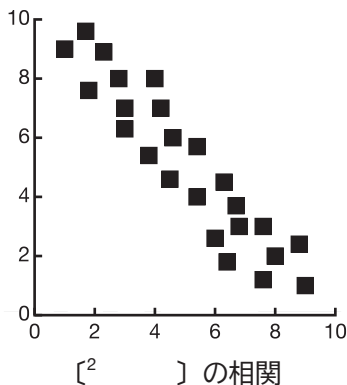
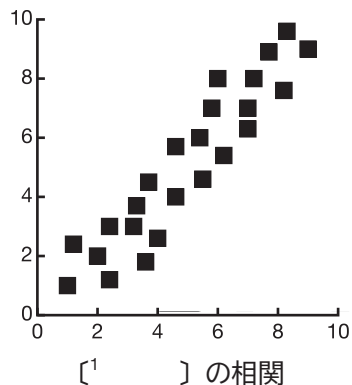
■ 回帰分析

相関関係

相関関係 = 2つの変量の間関係性

一方の変量が増加するにしたがってもう一方も増加する関係があるとき → **正の相関**

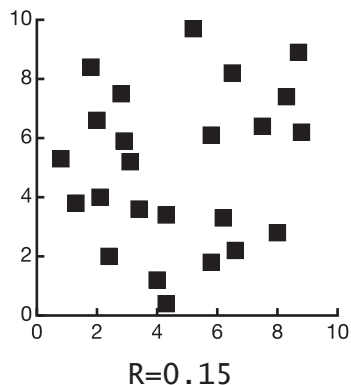
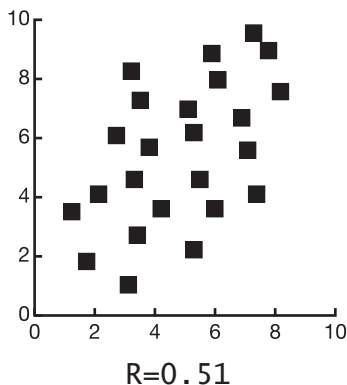
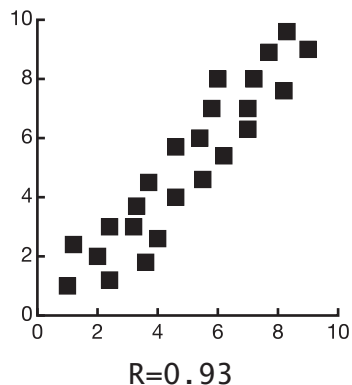
一方の変量が増加するにしたがってもう一方が減少する関係があるとき → **負の相関**



相関係数

相関係数Rは、-1から1までのいずれかの値をとる

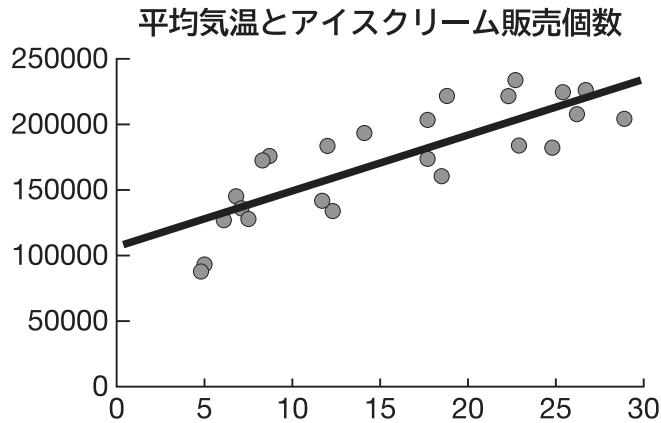
絶対値が1に近いほど相関が**強く**、0に近いほど相関が**弱い**



相関 [⁴] ←—————→ 相関 [⁵]

回帰分析

回帰分析= 変量間の相関を表す最も適した線を作成し、数式化する分析手法



※平均気温とアイスクリームの販売個数の相関をグラフで示すと↑のようなグラフに
 ※このグラフ中の点が実測値、直線で示したものが**回帰直線**（理論値）

回帰直線

上のグラフの直線のように、各データの真ん中を通る直線を**回帰直線**という
 →回帰直線を使うと、おおよその推測ができる（例：気温がもっと低くなったら 等）
 回帰直線の方程式は、下のような式で表される

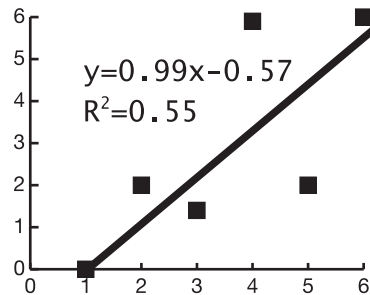
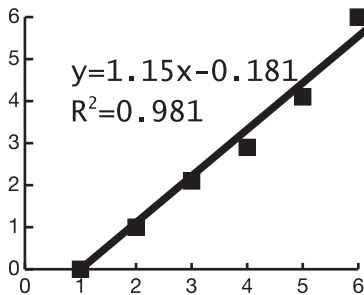
$$y=ax+b$$

※上のグラフの回帰直線の方程式は、次のように表される

$$y=4344x+105165$$


決定係数

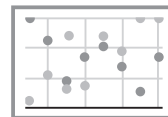
回帰直線と値がどれほど近いかの度合いを決定係数（ R^2 ）という
 決定係数は1に近いほど説明力が高く、0に近いほど説明力が低い



■ 相関グラフの作成方法

相関グラフの作成方法

- ①表の中にカーソルのある状態で「挿入→グラフ」を選ぶ
- ②グラフエディタで、グラフの種類を「**散布図**」にする
- ③[**X軸**] に横軸にしたい項目を、[**系列**] に縦軸にしたい項目を選ぶ
→複数系列がある場合は「:」から削除する



回帰直線の描画

- ④グラフエディタの [**カスタマイズ**] を選ぶ
- ⑤[**系列**] パネルを開ける
- ⑥[**トレンドライン**] にチェックを入れると回帰直線が表示される
- ⑦[**ラベル**] を「方程式」にすると回帰直線の方程式が表示される
- ⑧[**決定係数を表示する**] にチェックを入れると、決定係数 R^2 の値表示される

スクリーンショットの左側（設定）:

- 設定 **カスタマイズ** ④
- グラフの種類: 散布図 ②
- データ範囲: A1:C25
- X 軸: 123 平均気温 ③
- 系列: 123 販売個数 ③

スクリーンショットの右側（カスタマイズ）:

- 系列 ⑤: 販売個数
- 形式: 色: ●
- 塗りつぶしの色: 自動
- 塗りつぶしの透明度: 100%
- ポイントのサイズ: 7px
- ポイントの形: ●
- 軸: 左軸
- データポイントの書式を設定: 追加
- トレンドライン ⑥:
- ラベル: 方程式を使用 ⑦
- 決定係数を表示する ⑧:

相関係数を求める

相関係数を求めるには、CORREL()関数を使う

=CORREL(y軸のデータ,x軸のデータ)

※引数のx軸のデータとy軸のデータを入れる場所が通常とは逆なのに注意！

データ列の指定

データを列全体として指定する場合、「列名:列名」と書く

→例：B列を指定したい場合、「B:B」と表記

	A	B	C	D	E	F
1	年月	平均気温	販売個数		相関係数	
2	01/01	5.0	93,154			
3	02/01	6.8	145,152			
4	03/01	8.7	175,891			
5	04/01	14.1	193,424			
6	05/01	18.8	221,698			
7	06/01	22.3	221,5			

=CORREL(C:C, B:B)

相関係数の目安



表計算における関数

基本的にプログラミングにおける関数と同じようなものと思ってよい
関数の戻り値を表示させたいセルで関数を入力する
→関数は通常、下記のような形式で入力する

=関数名(引数1, 引数2, ...)

引数には、セル番地 (A1のような) や列番号 (A:Aのような) などが入る

どのような引数が必要かは関数によって異なる

ワークシート

問題1

右のデータは、年月ごとの平均気温とアイスクリームの販売個数の一覧です。相関グラフを作成し、以下の間に答えてください。

- ①相関関係はどうですか。
- ②回帰直線を描き、回帰直線の方程式を求めてください。
- ③回帰直線の決定係数 R^2 の値を求めてください。
- ④ある月の平均気温が 25.5°C と予想される時、その月のアイスクリームの販売個数はいくつくらいになると予想されますか。100の位を四捨五入し、1000個単位で答えてください。

年月	平均気温	販売個数
2017/01	5.0	93154
2017/02	6.8	145152
2017/03	8.7	175891
2017/04	14.1	193424
2017/05	18.8	221698
2017/06	22.3	221557
2017/07	25.4	224557
2017/08	26.2	207805
2017/09	22.9	183894
2017/10	17.7	173741
2017/11	12.3	133913
2017/12	7.1	135974
2018/01	4.8	87823
2018/02	6.1	126946
2018/03	8.3	172531
2018/04	12.0	183605
2018/05	17.7	203426
2018/06	22.7	233771
2018/07	26.7	226158
2018/08	28.9	204259
2018/09	24.8	182235
2018/10	18.5	160500
2018/11	11.7	141764
2018/12	7.5	127857

問題2

次のデータは、直近30日間の平均気温と肉まんの販売個数の一覧です。相関グラフを作成し、以下の間に答えてください。

- ①相関関係はどうですか。
- ②回帰直線を描き、回帰直線の方程式を求めてください。
- ③回帰直線の決定係数 R^2 の値を求めてください。
- ④予報では、2/15の最高気温が 5.0°C になると予想される時、肉まんはおよそいくつ売れると予想されますか。小数点以下を四捨五入して整数で答えってください。

日付	最高気温	販売個数
1/16	4.2	102
1/17	4.7	88
1/18	5.5	86
1/19	6.8	86
1/20	7.7	70
1/21	8.6	69
1/22	9.4	68
1/23	9.6	73
1/24	8.7	79
1/25	7.4	81
1/26	5.7	91
1/27	4.8	91
1/28	4.2	105
1/29	4.5	92
1/30	5.1	82
1/31	6.0	79
2/1	7.5	74
2/2	8.7	75
2/3	9.7	68
2/4	10.3	74
2/5	9.2	80
2/6	8.1	79
2/7	6.0	89
2/8	4.9	93
2/9	6.5	84
2/10	4.2	94
2/11	9.6	75
2/12	5.3	95
2/13	6.4	89
2/14	4.3	91

問題3

賃貸物件の家賃が何によって決まっているかを調べるために、駅からの徒歩時間、広さ、築年数と家賃額を表にまとめました。

築年数(年)	徒歩(分)	広さ(m ²)	家賃(円)
2	25	24.63	66000
2	10	20	70000
0	10	21.63	65000
14	8	19.51	69000
3	12	21.11	66000
6	8	17.32	65000
10	8	25.8	75000
9	4	22.68	69500
12	10	28.13	76000
3	13	26.99	69000
10	9	21.9	65000
10	9	19.87	67000
9	11	25.34	66000
9	7	20.81	61000
9	6	22.68	69500
5	5	21.73	74000
12	9	22.35	70000
12	13	23.72	74000
4	13	20.38	62000
11	15	26.67	69000
12	12	25.78	75000
3	5	25.01	77000
3	6	27.75	78000
19	10	18.3	64000
11	17	19.87	68000
10	18	29.84	74500
11	18	20.81	68500
9	20	26.08	68500
22	14	26	63000
17	19	30.43	59500
...
5	5	31.26	24000

(1)各観点ごとに相関グラフを作成してください。それぞれの相関関係はどうですか。

- ①徒歩と家賃
- ②築年数と家賃
- ③広さと家賃

(2)最終行に書いてある物件の家賃額が、とある事情で他と比べて極端に低いため、正確に相関関係を分析することができません。

最終行のデータを削除した上で、再度相関関係を考えてみてください。

- ①徒歩と家賃
- ②築年数と家賃
- ③広さと家賃

(3)(2)の結果から、家賃を決定づけている要因は何であると考えられますか。

振り返り

次の各観点が達成されていれば□を塗りつぶしましょう。

- 相関関係の意味を理解し、相関グラフの意味を読み取ることができるようになった
- 回帰直線の意味を理解し、回帰分析ができるようになった

今日の授業を受けて思ったこと、感じたこと、新たに学んだことなどを書いてください。

.....

.....

.....

.....

外部参照関数

Googleスプレッドシートには、別のスプレッドシートからデータを参照する関数が用意されています。データを参照ということは、データをコピーするわけではありません。したがって、元のデータが変更されれば、自動的に更新されます。

■ 外部参照関数

外部参照関数

別のスプレッドシートからデータを参照するには、IMPORTRANGE()関数を使う

```
=IMPORTRANGE("スプレッドシートID","データ範囲")
```

スプレッドシートID

スプレッドシートIDは、スプレッドシートのURLの/d/と/edit?に挟まれた箇所

```
https://docs.google.com/spreadsheets/d/[スプレッドシートID]/edit?...
```

データ範囲

データ範囲は、次のように指定する

```
シート名!開始列:終了列
```

	A	B	C	D	E
1	年齢	性別	満足度	推奨するか	
2	49	女性	とても不満	はい	
3	56	その他	とても不満	いいえ	
4	40	女性	とても不満	いいえ	
5	18	男性	とても満足	いいえ	
6	56	男性	普通	いいえ	
7	52	女性	とても不満	はい	

スプレッドシートIDが「abcdefg1234567」だったとする

上の表を参照するには、次のような数式となる

```
=IMPORTRANGE("abcdefg1234567","アンケート結果!A:D")
```


■ 問い合わせ関数

問い合わせ関数

別シートの表のデータを選択演算や射影演算するにはQUERY()関数を使う

```
=QUERY(データ範囲,"クエリ")
```

※データ範囲の指定方法は、左ページを参照

クエリ

データをどのように取得するかを指定するSQL風のテキスト

例) シート1のA:Dにアンケート結果があったとする

①満足度が「とても満足」のデータを取り出す

```
=QUERY(シート1!A:D,"select A,B,C,D where C='とても満足'")
```

※where句内で文字列を指定する場合、'で括る必要がある

②年齢が40歳未満のデータを取り出す

```
=QUERY(シート1!A:D,"select A,B,C,D where A < 40")
```

課題

ClassroomにあるスプレッドシートIDをコピーし、アンケート結果を外部参照させよう
上のクエリ2つを実行してみよう

スプレッドシートに共有設定を施し、URLを提出しよう

振り返り

次の各観点が達成されていれば□を塗りつぶしましょう。

- 外部参照関数を使えるようになった
- 問い合わせ関数を使えるようになった

今日の授業を受けて思ったこと、感じたこと、新たに学んだことなどを書いてください。

.....

.....

.....

アンケート分析

他の人のとったアンケート結果を組み合わせ、さまざまな分析をしてみよう。アンケートの回答はすべてメールアドレスでひも付けられており、同じ人物が別々の観点で答えた結果を参照することができます。そのため、相関関係等を分析することができます。

(教科書II：p.78 – p.79)

■ アンケート分析

アンケート結果の読み取り

Classroomで提供しているデータシートを開き、外部参照関数の「アクセスを許可」

- ① データを見て、相関関係のありそうな2つのデータを見つけ、仮説を立てる
- ② 問い合わせ関数を使い、①の2件のデータを並べる
- ③ ②で取り出したデータから散布図を作成し、データの相関関係を検証する
- ④ 検証結果から、相関について考察する

① 仮説を立てる

自分が考えた2つのデータの間にはどのような相関があるか（正の相関 or 負の相関）

なぜそのような相関があると考えたのかの理由を考える

2つのデータの見出しと列番号を下の表にメモをしよう

データの見出し	列番号

② データの取り出し

問い合わせ関数を使い、データを取り出す

```
=QUERY(データ!A:[最終列], "select [列番号1], [列番号2]")
```

※ [最終列] はClassroomを参照のこと

※ [列番号1] および [列番号2] は①の表に書いた番号を入力

③④ 検証&考察

②のデータからデータの散布図、相関係数等を求め、相関関係を検証する

検証結果から、相関関係があるかどうかを判定する

なぜそのような結果になったかについて考察する

※ 相関関係がある場合、因果関係があるかどうかも考えよう（疑似相関ではないか？）

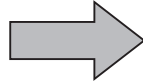
■ 相関関係と因果関係

相関関係

相関関係 = 二つの事柄が密接に関わりあい、一方が変化すれば他方も変化する関係



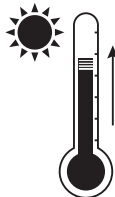
アイスクリームがたくさん売れる



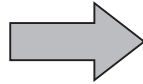
熱中症が増える

因果関係

因果関係 = 二つの事柄のうち、一方が原因で、もう一方が結果となる関係



気温が上がる

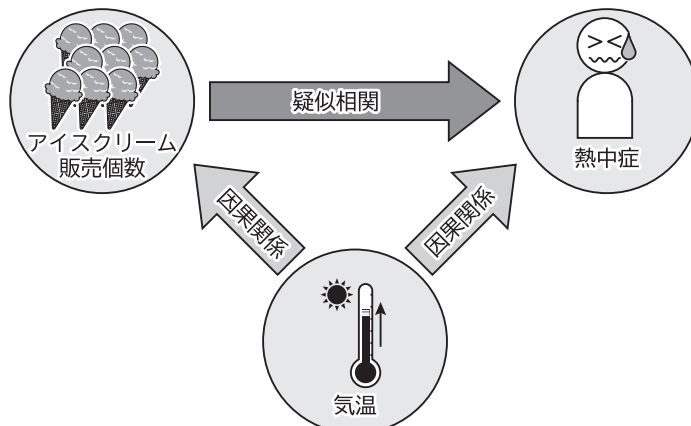


アイスクリームがたくさん売れる

疑似相関

疑似相関 = 二つの事柄に因果関係がないのに、あるように見えること

※疑似相関には、見えない要因が隠されている場合がある



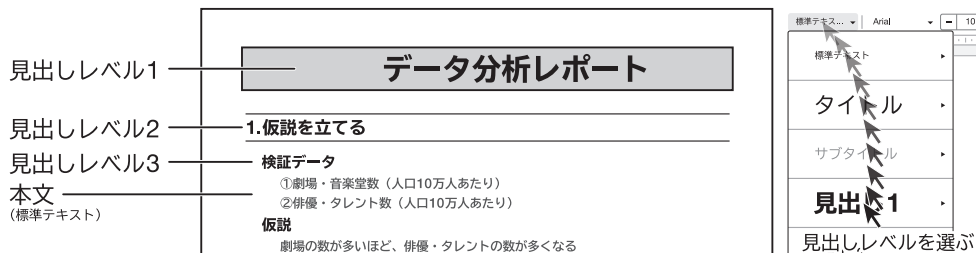
■ データ分析レポートの作成

レポートの作成

レポートのアウトライン化

レポートには見出しレベルと本文のレベルにスタイルの設定がされている

→見出しと本文のレベルを保持してレポートを書くこと（見出しはいじらない）



グラフの貼り付け

グラフを画像としてダウンロードし、ダウンロードした画像をレポートに貼り付ける

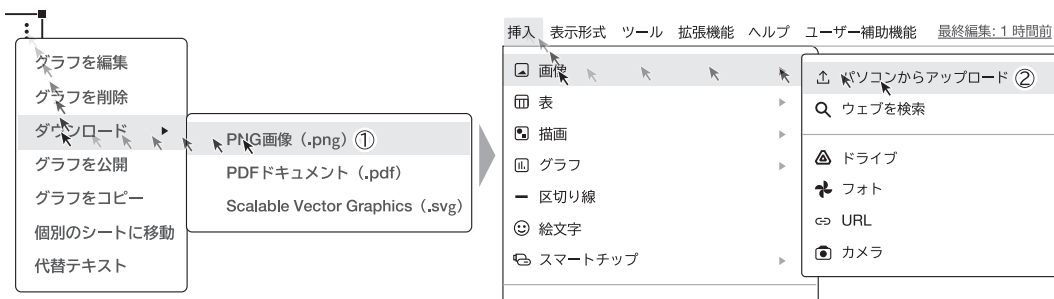
①グラフをダウンロードするには、

→グラフ右上の⋮から「ダウンロード▶PNG画像 (.png)」を選ぶ

②レポートに貼り付けるには、

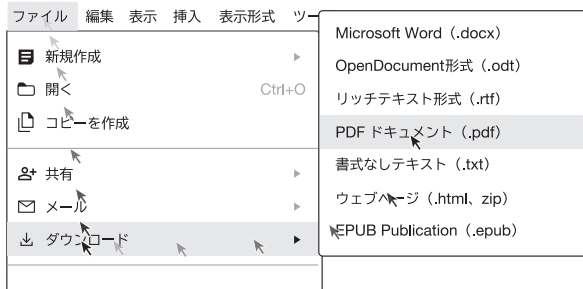
→「挿入→画像→パソコンからアップロード」を選ぶ

→グラフのファイル (chart.png) を選び [開く] を押す



PDFでのダウンロード

提出には、PDFでダウンロードしてPDFでダウンロードする
メニューの「ファイル→↓ダウンロード→PDF ドキュメント (.pdf)」を選ぶ



課題

データ分析レポートを書き、PDFファイルにして提出してください。

検証データに関しては、散布図、相関係数を求めるだけでなく、それぞれのデータの平均値、中央値、標準偏差も求めてください。(クロス集計を使おう)

評価に関して

すべての項目が書けており、かつ考察まできちんと書けているかどうかを評価

振り返り

次の各観点が達成されていれば□を塗りつぶしましょう。

- データの相関関係を調べることができるようになった
- 仮説、検証、考察の順でデータについて考えることができた

今日の授業を受けて思ったこと、感じたこと、新たに学んだことなどを書いてください。

.....

.....

.....

.....

章末問題

【問題1】

次の表は、各社員ごとの売上額の一覧表です。

Classroomにて配信したデータでクロス集計を行い、各問に答えてください。

社員名	部署	地域	売上額
佐藤	技術	北部	3235
鈴木	技術	南部	3737
高橋	営業	北部	3035
田中	人事	東部	2040
伊藤	人事	東部	4707
渡辺	技術	北部	4832
山本	人事	西部	3866
中村	財務	北部	2298
小林	営業	西部	3984
加藤	営業	北部	3657

(1) 売上額がもっとも高い部署はどこですか。

(2) 売上額がもっとも高い地域はどこですか。

(3) 社員1人あたりの売上額がもっとも高い地域はどこですか。

(4) もっとも売上額が高い地域・部署の組み合わせはどこですか。

【問題2】

次の表は、洗剤の価格と需要の関係を調査した結果です。

Classroomにて配信したデータでグラフ等を作成し、各問に答えてください。

価格	需要
462	216
252	365
349	260
231	412
356	286
377	250
247	348
302	334
463	202
269	329

(1) 相関関係はどうですか。

(2) 相関係数を求めてください。(小数点以下第2位まで)

(3) 回帰直線の方程式を書いてください。

(4) 価格を300円に設定したときの需要はいくらと予測されますか。

コラム～アンケートの弱点

■ アンケートのメリットとデメリット

アンケートのメリット

実施しやすい	調査票さえ用意すれば調査を実施できる
コストを抑えられる	対象者に一斉に実施できるため人件費や調査費を抑えられる
時間がかからない	多くの対象者に同時に実施できる
データを数値化できる	選択式の調査であれば容易に数値化できる

アンケートの弱点

アンケートには弱点もある

調査票は変更できない

アンケート調査は、一度開始したら調査票の変更ができない

→事前のアンケート設計が非常に重要

事実と異なる回答の可能性

アンケートでは回答に対する事実確認を行うことが困難

事実と異なる回答や虚偽の回答が含まれる可能性がある

回答の深堀が困難

選択肢形式の設問の場合、回答者は自分の考えに最も近いものを選択

→自分の意見を完全に表現できないと感ずることがある

→結果的に、得られる情報が限定的になる可能性が高い

自由記述式の設問を設けると、一般的に回答率が格段に落ちると言われている

回答の偏りの可能性

質問の言い回しや順番が回答に影響を与える可能性がある

回答者が特定の層に偏ると、調査結果が偏ってしまう可能性がある

アンケートは設計がいのち

アンケートは、目的を明確化させ、収集すべきデータを明らかにすることが最も大切

質問を簡潔かつ分かりやすくし、回答の幅を持たせ、漏れがないように設計する必要

