

## データサイエンス講座①

### ★ テーマを確認(1 min)

- データの分析で求めてきた、分散や相関係数などの統計量をコンピュータで計算しよう



### ★ 統計量の確認(10 min)

- ワークシート **1** に取り組む。
- 必要あれば定義はスライドで確認。

#### 《統計量の復習》

データの分析で学んだことを思い出し、

問題1に取り組もう！

平均値  $\bar{x} = \frac{1}{n}(x_1 + x_2 + \dots + x_n)$

分散  $s_x^2 = \frac{1}{n}\{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \dots + (x_n - \bar{x})^2\}$

標準偏差  $s_x = \sqrt{s_x^2}$

共分散  $s_{xy} = \frac{1}{n}\{(x_1 - \bar{x})(y_1 - \bar{y}) + \dots + (x_n - \bar{x})(y_n - \bar{y})\}$

### ★ **1**の答え合わせ(5 min)

- 表はスライドにあり
- 求め方や答の確認

**1**

番号	x	y	x - $\bar{x}$	y - $\bar{y}$	(x - $\bar{x}$ ) <sup>2</sup>	(y - $\bar{y}$ ) <sup>2</sup>	(x - $\bar{x}$ )(y - $\bar{y}$ )
①	36	27	-1	-6	1	36	6
②	42	39	5	6	25	36	30
③	35	35	-2	2	4	4	-4
④	33	25	-4	-8	16	64	32
⑤	38	41	1	8	1	64	8
⑥	32	23	-5	-10	25	100	50
⑦	39	43	2	10	4	100	20
⑧	40	31	3	-2	9	4	-6
⑨	34	29	-3	-4	9	16	12
⑩	41	37	4	4	16	16	16
計	370	330	0	0	110	440	164

★ スプレッドシートで統計量の計算(20+9 min)

【ワーク①】(20 min?)

- 各自生徒クロムブックでクラスルーム/探究基礎/データサイエンス① を開くよう指示。
- ペア学習で助け合い
- データが大きいと手計算が大変である、アンケートの集計・加工に利用できる、など、コンピュータを用いるモチベーションに触れておく。
- 吹き出し箇所の数式について説明する。
- ある程度作業させてから、オートフィル(Ctrl+引張る) 半角で入力 などをコメント
- さらにしてから、相関係数が “=correl(C2:C48,D2:D48)”で求まることや、その他よく手間取っていたポイントについて触れる

《表計算アプリで統計量を計算しよう》

クラスルーム/探究基礎/データサイエンス①

色付きセルに数式等を入れて、統計量を求めよう

=C2 - \$L\$5  
(C2セルの値) - (L5セルの値)

=E2^2  
E2セルの値の2乗

=MAX(C2:C48)  
C2セルからC48セルまでの合計値

番号	地域	x: 年間収入 【千円】	y: 金利費 (全世帯) 【円】	xの偏差 (xの値-xの平均)	yの偏差 (yの値-yの平均)	xの偏差× yの偏差	xの2乗 (xの値の2乗)	yの2乗 (yの値の2乗)	xの共分散 xとyの平均値	yの共分散 xとyの相関係数
1	北海道	4488	63638							
2	青森県	4952	65700	=C2-\$L\$5	=E2-\$E\$5		=E2^2			
3	岩手県	5282	68745							
4	宮城県	5702	67836							
5	秋田県	5274	66067							
6	山形県	5895	66362							
7	福島県	5357	65103							
8	茨城県	5965	65872							
47	沖縄県	4215	50801							
計		=SUM(C2:C48)								

【ワーク②】(9 min?)

- 時間調整用なので、ワーク①の進み次第で割愛可(ここまでいかない気がする…)
- zかwのデータから1つ選び、xとの相関関係を計算する
- ペアで別のものを調べ、共有するとよい

★ 作業がキリの良いところでおしまい

## データサイエンス講座②

### ★ テーマを確認(1 min)

- データの分析ではあらかじめ与えられることが多かったヒストグラムや度数分布表、散布図をコンピュータで作成する方法を学ぶ。
- クラスルーム/探究基礎/データサイエンス講座②を開く



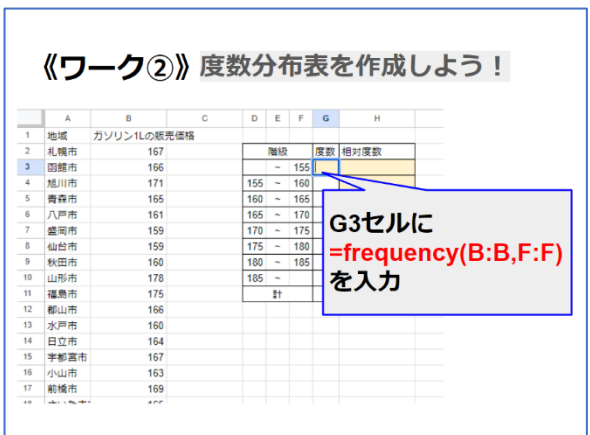
### ★ 1 ワーク① ヒストグラムの作成 (10 min)

- 列 A を選択しながら挿入→グラフでヒストグラムが作成されることを説明し、実際に作成させる。
- 画面右側のグラフエディタ内  
カスタマイズ→ヒストグラム→バケットサイズ  
  - これが階級の幅を表す。
  - カスタマイズ、グラフと軸のタイトル、横軸・縦軸をいろいろいじらせて、グラフを完成させる



### ★ 2 ワーク② 度数分布表の作成(15 min)

- G2セルに”=frequency(B:B,F:F)”と入力することで B 列のデータで F 列を区切りにした度数分布が作れることを説明。
- 色付きセルに適切な数式を入力し、度数分布表を完成させる。
- 早く終わった生徒にはヒストグラムも作成させる。



★ **3**ワーク③ 散布図の作成(9 min)

- ・ 散布図作成の手順を説明
- ・ 軸の最大値・最小値をいろいろ調整させ、見やすく整理

《ワーク③》 散布図を作成しよう！

列B:Cを選択しながら、挿入 → グラフ

グラフエディタ → 散布図

★ 回帰直線の描画(10 min)

- ・ 系列→トレンドラインで回帰直線が引けること
- ・ 散布図でできるだけ合うように引いた直線であること  
 くらいのざっくりした説明でよいかと。
- ・ さらに、トレンドライン下部のラベル（方程式を使用）で回帰直線の方程式を表示できること  
 $(y=0.391*x+10)$  やこれを用いて、握力 43kg の場合のハンドボール投げの予測ができることに触れる（時間があれば）

《ワーク③》 回帰直線を描画しよう！

回帰直線を引くには... 系列 →トレンドライン

方程式を使用

## データサイエンス講座③

### ★ テーマを確認(1 min)

- これまで2回でデータを整理する方法を学んだ。今回は整理したデータから実際に分析をしてみよう。

### ★ 題材の理解・仮説の説明(4min)

- 今回は統計データから予め箱ひげ図や散布図に整理してあり、ここからどのようなことが分かるか考察する。
- 仮説の説明「観察された事実をもとに推論した、自分なりの「答え」となる命題のこと」

### ★ 【ワーク①】仮説の例の説明と仮説を立てる(10min)

- 仮説の例で説明(命題の形で表現されること、この場合食事時間をもとに分析ができることなど)
- (1)に取り組む
- (2)表から分析可能な仮説を自分で立ててみる  
→4人くらいのグループで共有

### ★ 【ワーク②】箱ひげ図から読み取り(15min)

- (1)~(3)に取り組む
- (2)・(3)はグループで共有する

### ★ 【ワーク③】散布図から読み取り(15min)

- 紙のデータだけでは読み取れることに限界があるが、電子データ【スプレッドシート：データサイエンス講座③】で相関係数や回帰直線をかいて考察するとよい。
- 考察後、グループで共有する



### 《今回のテーマ》

都道府県別生活時間の統計データから、  
仮説を立てて分析しよう

仮説とは・・・

観察された事実をもとに推論した、  
自分なりの「答え」となる命題のこと

※ 探究の結果、  
正しい 間違い 不完全である  
といった判断をする

### 《ワーク①》

#### 仮説の例

若年層でスマートフォン・パソコンなどの使用時間が長いグループは、使用時間が短いグループよりも食事の時間が短くなる傾向がある。

それぞれの表の食事時間を  
比較することで分析できる

表1-1-1 スマートフォン・パソコンなどの使用時間と都道府県別平均生活時間

都道府県	睡眠時間 (分)	身の回り作業時間 (分)	食事時間 (分)	通学時間 (分)	授業時間 (分)	通勤・帰宅時間 (分)
北海道	429	79	60	66	66	4
東京都	411	77	52	96	69	13
東京都	409	79	52	96	69	13

表1-1-2 スマートフォン・パソコンなどの使用時間と都道府県別平均生活時間

都道府県	睡眠時間 (分)	身の回り作業時間 (分)	食事時間 (分)	通学時間 (分)	授業時間 (分)	通勤・帰宅時間 (分)
北海道	436	77	65	63	61	6
東京都	404	78	58	99	69	13
東京都	443	78	62	63	63	6

## データサイエンス講座④

### ★ テーマを確認(1 min)

- ・ 数Iで学習した仮説検定の考え方について理解を深めよう、くらいな感じで。

### ★ 仮説検定の考え方を実際の問題で復習 (10 min)

- ・ まずは各自で解かせる。
- ・ スライドで解説。
- ・ 実際の場面では、あらかじめコイン投げの表は準備されていない  
→ コイン投げ実験も自分で準備する必要があるが大変  
→ コンピュータシミュレーションで手軽に実験できる

### ★ コイン投げシミュレーションで遊ぶ(15min)

- ・ クラスルームのコイン投げシミュレーション URL からアクセス
- ・ 吹き出し3か所を中心に、使い方を説明
- ・ 2について自分で実験回数や確率の基準を設定して、シミュレーションを実施して取り組む。
- ・ 早く終わった生徒は3も (終わらなくてもよい)
- ・ その他いろいろじって遊ばせる。
- ・ コイン投げの回数が多くなると、正規分布に近づいていくことを実感させる。  
下部に正規分布曲線を描画するボタンあり(webサイトを投影して実際に見せてもよい)。



#### 《仮説検定の考え方》

1 タイムがよくなるのは全くの偶然で起こる、と仮定する . . . ①

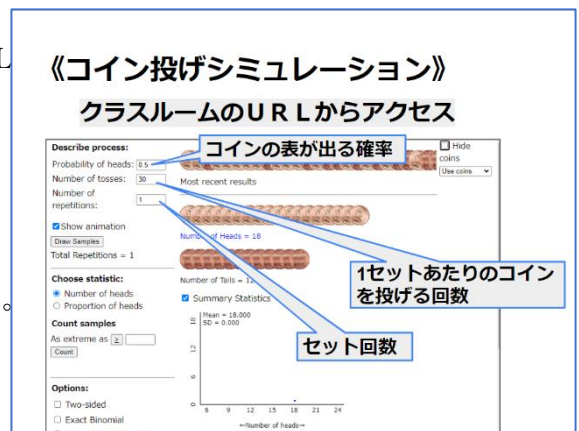
公正なコインを30回投げて、表が20回以上出る相対度数は

$$\frac{7+6+0+1}{300} = 0.046\text{.....}$$

表の回数	8	20	21	22	23	計
度数	3	7	6	0	1	300

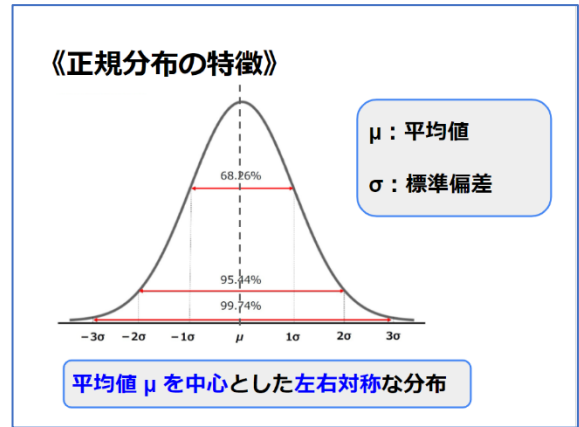
であり、0.05 より小さい

よって、仮定①が正しいとは判断できないから、タイムがよくなると判断してよい



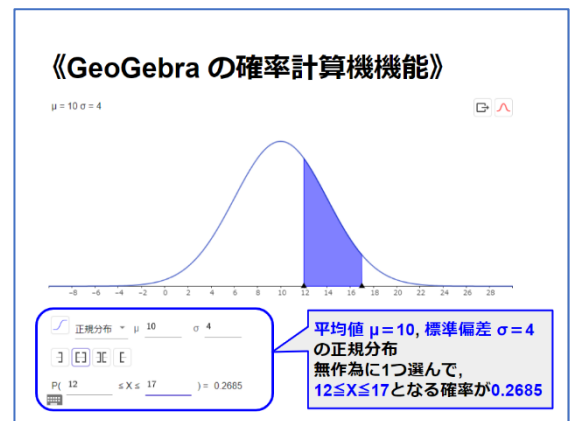
★ 正規分布の説明(5 min)

- ・ 平均値  $\mu$  を中心とした左右対称な分布、標準偏差  $\sigma$  の幅を区切りとした分布の割合、および、確率の基準(0.05, 0.01)の根拠をざっくりと説明。



★ GeoGebra の確率計算機能を用いて4に取り組む(14 min)

- ・ クラスルームの URL からアクセスできる。
- ・ スライド枠囲みの部分の説明  
(実際の GeoGebra 画面を投影してもよい)。
- ・ 4に取り組む。(1)は解説,  
(2)(3)はいろいろいじらせて任せてもよい。



## データサイエンス講座⑤

### ★ テーマを確認(1 min)

- ・ さっそく2枚目のスライドへ

### ★ 集団を比較する場面の例(14 min)

- ・ スライドで例示
- ・ 【ワーク①】でほかの例も考えてみる
- ・ 【ワーク②】例1～3の違いを考察する
  - ◇ 何と何を比較しているか、対象は同じか・違うか、差があるか・増減まで考慮しているかなど考察の観点を与えてもよい。

### ★ 仮説検定の流れ(10 min)

- ・ 例1を題材に仮説検定の流れを確認
- ・ 用語の説明（帰無仮説、p値、有意水準）

### ★ 比較の分類についてまとめ(10 min)

- ・ 用語の確認  
(1群2群, 対応ありなし, 両側検定片側検定)
- ・ 種類によって検定の手法が変わることを伝え、動画見せる(約5分)
- ・ (スプレッドシート⑤をやりたければやる)

### ★ アンケート(10 min)

- ・ クラスルームからフォームアンケートへ  
(スプレッドシート⑤をやりたければやる)



#### 《集団を比較する場面》

(例1) 100g入りのお菓子から、10袋抽出したところ、平均101.7gであった。

(例2) あるテストで、A組の平均値60.2点、B組の平均値59.4点であった。

(例3) 新薬の効果を検査するため、100人ずつ計200人に新薬または偽薬を投与して効果を調べたところ、表のとおりとなった。

	効能あり	効能なし
新薬	78	32
偽薬	51	39

#### 《仮説検定の流れ》

(例1) 100g入りのお菓子から、10袋抽出したところ、平均101.7gであった。

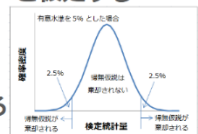
100gとなるように生産されている、と仮定する

この仮定の下で、

今回以上の差が生じる確率を調べる

この確率と基準の確率を比較し、

仮定が正しくないと判断できるか考察する



#### 《2つの集団の比較の種類》

- ① 全体(母集団)との比較なのか **1群**  
ある標本と別の標本の比較なのか **2群**
- ② 異なる対象への調査なのか **対応なし**  
同一の対象への調査なのか **対応あり**
- ③ 差があるかどうかに着目しているのか **両側検定**  
増加(減少)まで着目しているのか **片側検定**