

令和5年度 1年探究基礎

データサイエンス講座

H 番 氏名

目的 具体的なデータを題材に数値データの取り扱いに関する一連の手法を
学び，課題研究の充実を図る

講 座 内 容

講座① スプレッドシートを活用して統計量を求める

講座② スプレッドシートを活用してグラフを作成する

講座③ データから仮説を立てて分析する

講座④ 仮説検定の考え方と正規分布

講座⑤ 集団のデータに差があるか比較する

《講座① スプレッドシートを活用して統計量を求める》

1 下の表は、10 人の生徒の右手の握力と左手の握力を測定した結果である。

右手の握力 (x) と左手の握力 (y) の間には、どのような相関があると考えられるか、

相関係数 r をもとに調べてみよう。

	①	②	③	④	⑤	⑥	⑦	⑧	⑨	⑩
右手の握力 (kg)	36	42	35	33	38	32	39	40	34	41
左手の握力 (kg)	27	39	35	25	41	23	43	31	29	37

(1) 表を利用して、右手の握力 (x) と左手の握力 (y) について、次の統計量を計算しよう。

【最大値・最小値・平均値・分散・標準偏差・共分散・相関係数】

ただし、必要があれば、 $\sqrt{11} = 3.317$ とし小数第 3 位を四捨五入せよ。

番号	x	y	$x - \bar{x}$	$y - \bar{y}$	$(x - \bar{x})^2$	$(y - \bar{y})^2$	$(x - \bar{x})(y - \bar{y})$
①	36	27					
②	42	39					
③	35	35					
④	33	25					
⑤	38	41					
⑥	32	23					
⑦	39	43					
⑧	40	31					
⑨	34	29					
⑩	41	37					
計							

x の最大値	x の最小値	y の最大値	y の最小値
x の分散 $s_x^2 =$	x の標準偏差 $s_x =$	y の分散 $s_y^2 =$	y の標準偏差 $s_y =$
x と y の共分散 $s_{xy} =$		x と y の相関係数 $r =$	

(2) 右手の握力 (x) と左手の握力 (y) の間にはどのような相関があると考えられるか、説明せよ。

《表計算アプリケーションを活用して統計量を計算してみよう》

大きなデータになると、手計算で統計量を計算するのは困難になる。このようなときに活躍するのが表計算アプリケーションである。表計算アプリケーションの機能を使って、実際に統計量を計算してみよう。

2 クラスルーム/探究基礎/データサイエンス講座① を開いて、統計量を計算してみよう。

【ワーク①】色付きのセルに適切な数式を入れて表を完成させ、統計量を求めよ。

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	番号	地域	x: 年間収入 【千円】	y: 食料費 (全世帯) 【円】	xの偏差	yの偏差	(xの偏差) ²	(yの偏差) ²	(xの偏差)× (yの偏差)			x: 年間収入 【千円】	y: 食料費 (全世帯) 【円】	
2	1	北海道	4488	63638	=C2-\$L\$4	=E2*2					最大値	=max(C2:C48)		
3	2	青森県	4952	65700							最小値			
4	3	岩手県	5282	68745							平均値	5405.55319		
5	4	宮城県	5702	67836							分散			
6	5	秋田県	5274	66097							標準偏差			
7	6	山形県	5856	68382										
8	7	福島県	5397	65103							xとyの共分散		xとyの相関係数	
9	8	茨城県	5956	65872										
48	47	沖縄県	4215	50801										
49		計	=SUM(C2:C48)											

★ 入力のヒント

数式はすべて半角英数字で入力

算術演算子	意味
+	足し算
-	引き算
*	掛け算
/	割り算
^	累乗

関数	意味
MAX	最大値
MIN	最小値
SUM	合計値
AVERAGE	平均値
SQRT	平方根

L5セルを”L5”とかくのと,”\$L\$5”とかくのは、何か違いがあるのかな？

=MAX(C2:C48) : C2セルからC48セルまでの平均値

このほかにも、分散や標準偏差、相関係数を求める関数はあるのかな？

【ワーク②】

都道府県別の z : 自動車所有数量 (二人以上世帯) [台], w : ボランティア活動・社会参加活動の平均時間 (15歳以上男) [分] のデータからどちらかを選び, x: 年間収入 [千円] との相関関係を調べ, 説明せよ。

x と	の相関係数	r =

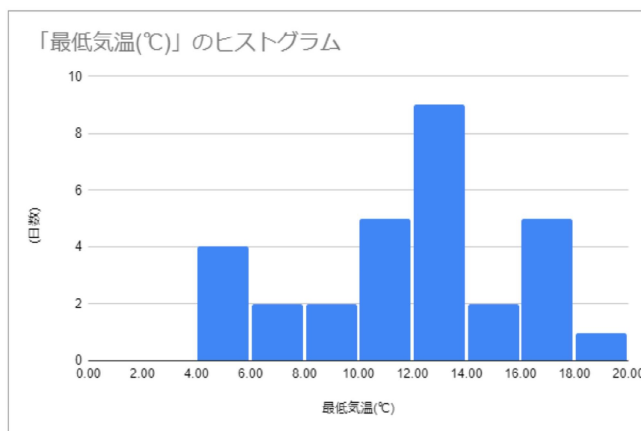
《講座② スプレッドシートを活用してグラフを作成する》

《表計算アプリケーションを活用してデータをグラフ化しよう》

クラスルーム/探究基礎/データサイエンス講座② を開いて、グラフ機能を活用したデータを可視化しよう。

- 1 【ワーク①】 次のデータはある都市のある月の日ごとの最低気温である。下記のようなヒストグラムを作成しよう。

7.1	10.7	8.9	7.5	11.0	12.6	17.0	18.6	16.5	13.9
10.1	12.6	14.1	17.6	14.0	11.7	16.9	16.3	13.7	13.5
12.2	13.3	11.4	12.5	12.2	4.9	5.0	8.6	5.6	4.4 (単位は℃)



★ 作成のヒント

- 画面上部 **挿入** → グラフ
- 画面右側 **グラフエディタ**
 - **カスタマイズ** → ヒストグラム内 バケットサイズ：階級の幅のこと
 - **グラフと軸のタイトル**
 - **横軸** ・ **縦軸**

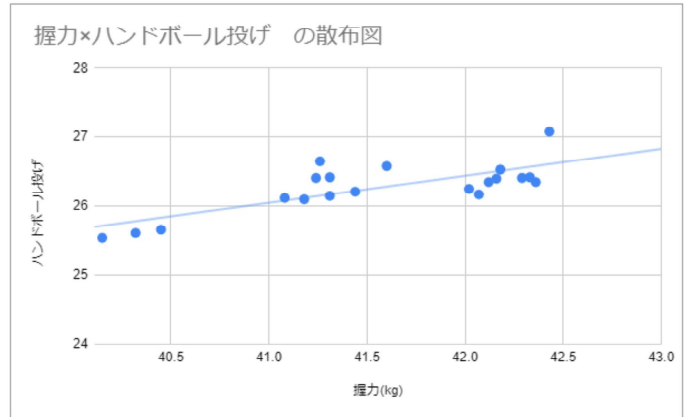
- 2 【ワーク②】 2023年1月における都市別のガソリン1Lあたりの小売価格のデータ(A:B列)がある。色付きセルに数式を入力することにより、階級の幅 5(円)の度数分布表を作成しよう。また、このデータのヒストグラムを作成せよ。

階級	度数	相対度数
~ 155		
155 ~ 160		
160 ~ 165		
165 ~ 170		
170 ~ 175		
175 ~ 180		
180 ~ 185		
185 ~		
計		

★ 作成のヒント

- G3セルに ”=frequency(B:B,F:F)” を入力

- 3 【ワーク③】体力運動能力調査(政府統計の総合窓口)による，体力テストにおける握力とハンドボール投げのデータ(B:C 列)から，図のような散布図を作成しよう。さらに，その散布図に回帰直線を描画しよう。



★ 作成のヒント

- ・ 軸で表示範囲の最大値・最小値を調整しよう。
- ・ 回帰直線は 系列→トレンドラインで描画

さらに，ラベル（方程式を使用）で回帰直線の方程式を読み取ろう。

回帰直線の方程式

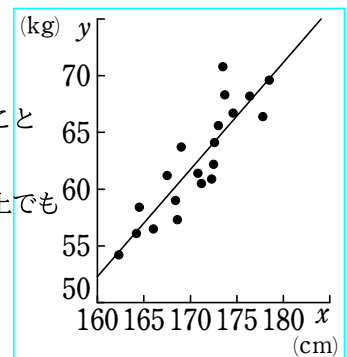
回帰直線をもとに，握力43kgの人のハンドボール投げの記録を推測しよう(小数第3位を四捨五入して答えよ)。

m

回帰直線とは...

散布図で点の配列に「できるだけ合うように引いた直線」のことである

回帰分析は自然科学や経済学や社会学など幅広い分野を学ぶ上でも重要な手法である。



《講座③ データから仮説を立てて分析する》

1 次の文章を読み，【ワーク①】～【ワーク③】に取り組もう。

次の表 1 は，国が実施した生活時間の実態に関する統計調査をもとに，15 歳以上19 歳以下の若年層について，都道府県別に平日 1 日の中で各生活行動に費やした時間（分）の平均値を，スマートフォン・パソコンなどの使用時間をもとにグループに分けてまとめたものの一部である。ここでは，1 日のスマートフォン・パソコンなどの使用時間が 1 時間未満の人を表 1－A，3 時間以上 6 時間未満の人を表 1－Bとしている。

表 1－A：スマートフォン・パソコンなどの使用時間が

1 時間未満の人の生活行動時間に関する都道府県別平均値

都道府県	睡眠 (分)	身の回りの 用事 (分)	食事 (分)	通学 (分)	学業 (分)	趣味・娯楽 (分)
北海道	439	74	79	60	465	8
青森県	411	74	73	98	480	13
茨城県	407	61	80	79	552	11
栃木県	433	76	113	50	445	57

表 1－B：スマートフォン・パソコンなどの使用時間が

3 時間以上 6 時間未満の人の生活行動時間に関する都道府県別平均値

都道府県	睡眠 (分)	身の回りの 用事 (分)	食事 (分)	通学 (分)	学業 (分)	趣味・娯楽 (分)
北海道	436	74	88	63	411	64
青森県	461	57	83	55	269	44
茨城県	443	80	81	82	423	63
栃木県	386	120	79	77	504	33

(出典：総務省統計局の平成 28 年社会生活基本調査により作成)

花子さんたちは，表 1－A をスマートフォン・パソコンなどの使用時間が短いグループ，表 1－B をスマートフォン・パソコンなどの使用時間が長いグループと設定し，これらのデータから，スマートフォン・パソコンなどの使用時間と生活行動に費やす時間の関係について分析してみることにした。ただし，表 1－A，表 1－B において一か所でも項目のデータに欠損値がある場合は，それらの都道府県を除外したものを全体として考える。なお，以下において，データの範囲については，外れ値も含めて考えるものとする。

【ワーク①】 《データから分析できる仮説を立てよう》

花子さんたちは、これらのデータからいくつかの仮説を考えることにした。

(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。

※ この仮説は表1－A, Bの食事時間を比較することで分析することができる。

(1) 次の2つの仮説①②について、表1－A, 表1－Bのデータだけでは分析できるか、できる場合には○、できない場合には×を記入せよ。

①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に注目すると、スマートフォン・パソコンなどを朝よりも夜に長く使っている 傾向がある。

②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に注目すると、学業の時間が長い都道府県は趣味・娯楽の時間が短くなる傾向がある。

①		②	
---	--	---	--

(2) 上記以外で表1－A, 表1－Bのデータだけでは分析できる仮説を立てて、さらにグループで共有しよう。

自分で立てた仮説

グループで共有した仮説

【ワーク②】 《箱ひげ図から分析しよう》

花子さんたちは表 1－A、表 1－B のデータから睡眠の時間と学業の時間に注目し、それぞれを図 1 と図 2 の箱ひげ図（外れ値は○で表記）にまとめた。

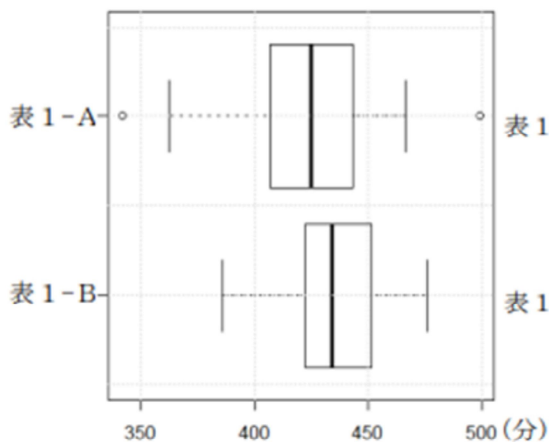


図 1 睡眠の時間の分布

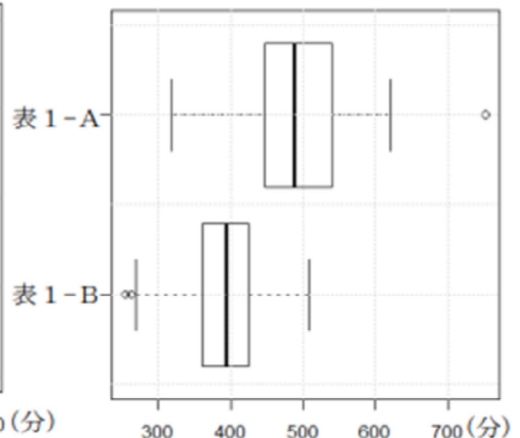


図 2 学業の時間の分布

(1) 図 1, 2 の箱ひげ図から読み取れるものとして最も適当なものを, ①～③から一つ選べ。

- ① 睡眠の時間が 420 分以上である都道府県の数を見たとき, 表 1－A の方が表 1－B よりも多い。
- ② 学業の時間が 550 分以上の都道府県は, 表 1－A においては全体の半数以上あり, 表 1－B においては一つもない。
- ③ 学業の時間が 450 分未満の都道府県は, 表 1－B においては全体の 75% 以上であり, 表 1－A においては 50% 未満である。

(2) (1)以外で図 1, 2 の箱ひげ図からどのようなことが読み取れるだろうか。

また, その内容をグループで共有しよう。

箱ひげ図から読み取れること

グループで共有した内容

(3) 花子さんたちは、スマートフォン・パソコンなどの使用時間の長さの違いが、睡眠の時間と学業の時間のどちらに大きく影響しているかについて調べることにした。そのために、都道府県ごとに睡眠の時間と学業の時間のそれぞれにおいて、表1－Aの値から表1－Bの値を引いた差について考え、その結果を次の図3の箱ひげ図（外れ値は○で表記）で表した。

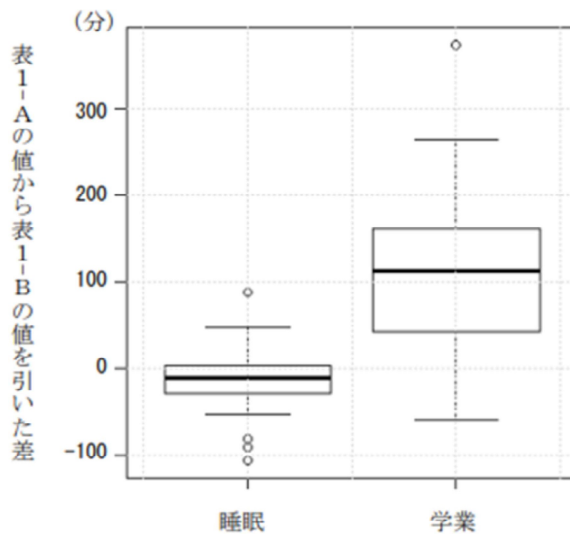


図3 生活行動時間の差

（図3から読み取ることができる内容の例） 学業の時間の差が正の値になっている都道府県の若年層は、スマートフォン・パソコンなどの使用時間が短いグループの方が、学業の時間が長い傾向にある。

図3から読み取ることができる内容を2つ以上あげ、さらにグループで共有しよう。

図3から読み取ることができる内容

・

。

・

グループで共有した内容

【ワーク③】 花子さんたちは、表1-Aについて、睡眠の時間と学業の時間の関連を調べることとした。次の図4は、表1-Aについて学業の時間と睡眠の時間を散布図で表したものである。ただし、2個の点が重なって区別できない場合は□で示している。

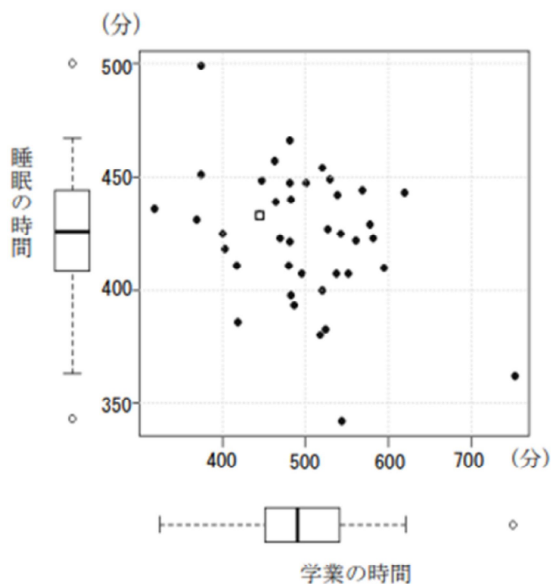


図4 表1-Aの学業の時間と睡眠の時間の散布図

また、この散布図のスプレッドシートはクラスルーム/探究基礎/データサイエンス講座③にある。このデータからどのようなことが分かるだろうか、また、その内容をグループで共有しよう。

このデータから分かること

グループで共有した内容

M E M O

《講座④ 仮説検定の考え方と正規分布》

- 1 あるスポーツメーカーが、新しいシューズ S を開発した。無作為に選んだ高校生 30 人が、1 回目はシューズ S 以外のシューズを履いて、2 回目はシューズ S を履いてそれぞれ 50 m 走のタイムを計測したところ、20 人が 1 回目より 2 回目のタイムがよかった。このデータから、シューズ S を履くことでタイムがよくなると判断してよいか。仮説検定の考え方をを用い、基準となる確率を 0.05 として考察せよ。ただし、公正なコインを 30 回投げて表の出た回数を記録する実験を 300 セット行ったら、次の表のようになったとし、この結果を用いよ。また、1 回目の計測は 2 回目の計測に影響を及ぼさないものとする。

表の回数	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	計
度数	3	7	10	13	19	33	41	45	44	36	22	13	7	6	0	1	300

《公正なコイン投げ実験をコンピュータシミュレーションで効率よく行ってみよう》

★ コイン投げシミュレーション

コインの表が出る確率

Describe process:

Probability of heads: 0.5

Number of tosses: 30

Number of repetitions: 1

☒ Show animation

Total Repetitions = 1

Choose statistic:

☒ Number of heads

☐ Proportion of heads

Count samples

As extreme as

Options:

☐ Two-sided

☐ Exact Binomial

☐ Normal Approximation

Hide coins

Use coins

Most recent results

Number of Heads = 19

Number of Tails = 11

Summary Statistics

1セットあたりのコインを投げる回数

セット回数

Number of heads

- 2 あるコインを 80 回投げたところ、表が 48 回出た。このとき、このコインは表が出やすいと判断してよいか。コイン投げシミュレーションを行い、その結果用いて判断せよ。なお、確率の基準も自分で設定すること。

解答

公正なコインを1セット_____回投げる実験を_____回 行うシミュレーションを

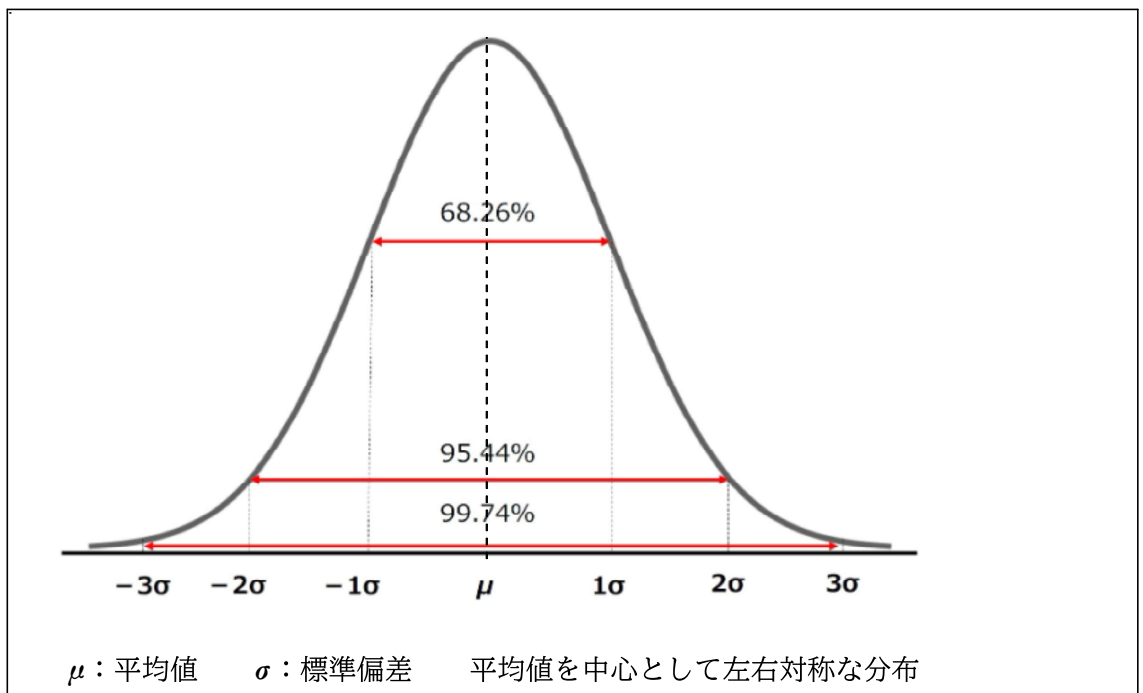
行ったところ、表が_____ 回以上出る場合の相対度数は、

これは基準となる確率_____より_____から、

このコインは_____

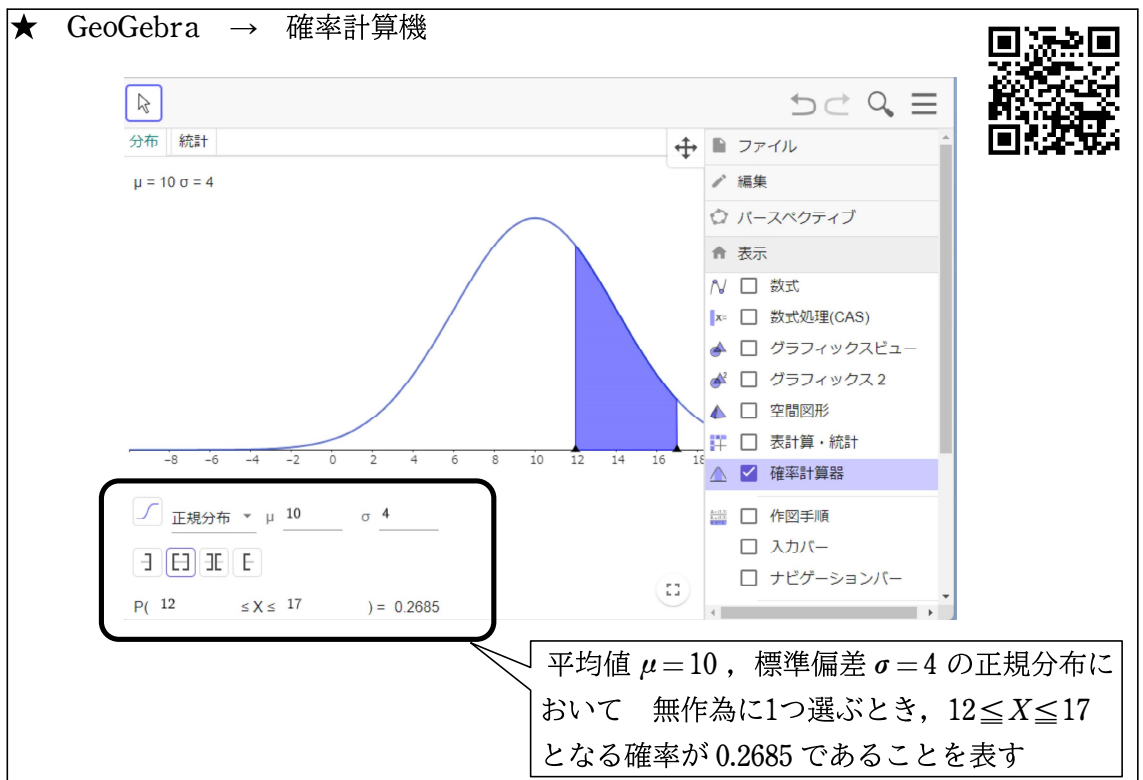
- 3 ある都市には、ランドマークともいえるタワー T がある。しかし、老朽化が進んだため、取り壊すかどうかの住民投票が行われた。そのときは全住民の $\frac{1}{8}$ が取り壊しに賛成した。その 2 年後、今度は住民 100 人に対してアンケートを実施したところ、19 人が賛成と答えた。このとき、タワー T の取り壊しに賛成する人が増えたと判断してよいか。コイン投げシミュレーションを行い、その結果用いて判断せよ。なお、確率の基準も自分で設定すること。

《正規分布の特徴》



《GeoGebraの確率計算機能》

★ GeoGebra → 確率計算機



4 ある試験での成績の結果は、平均点 64 点、標準偏差 14 点であった。得点の分布を正規分布とみなすとき、次の問いに答えよ。ただし、(1)(2)は小数第 2 位を四捨五入するものとする。

- (1) 70 点以上 90 点以下の人は約何%いるか。
- (2) 合格点を 50 点とすると、約何%が合格することになるか。
- (3) この試験で成績上位3%以内に入るには何点以上をとる必要があるか。

《講座⑤ 集団のデータに差があるか比較する》

《集団を比較する場面》

（例１）ある工場では1袋100g入りのお菓子を生産している。このお菓子を10袋抽出して質量を検査したところ平均値が101.7gであった。この工場では100gのお菓子を出荷していると判断してよいだろうか。

（例２）ある学校で行った試験の点数について、A組の平均点は60.2点、B組の平均点は59.4点であった。A組とB組で試験の結果に差があると判断してよいだろうか。

（例３）新薬が開発されたときには、その薬が本当に効くかどうかを次のようにテストする。
新薬または偽薬（プラシボ）のどちらかを被験者に教えずに与えて、効能があったかどうか調べる。
ある新薬で100人ずつ計200人の被験者を対象にこのテストを行ったところ、表のとおりとなった。この新薬は有効であると判断してよいのだろうか。

	効能あり	効能なし
新薬	78	32
偽薬	51	39

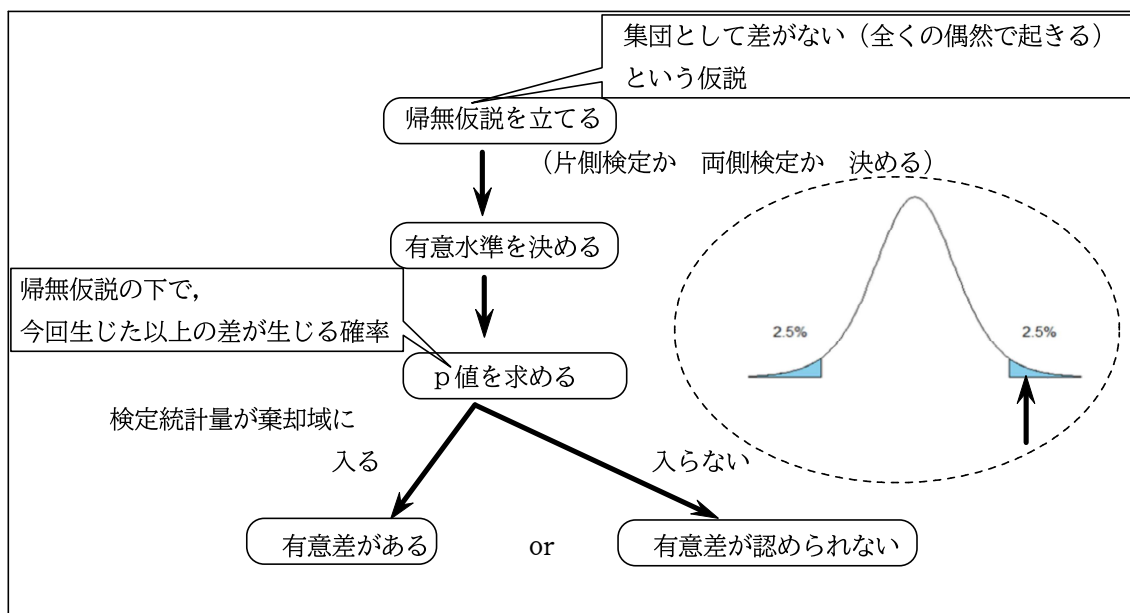
【ワーク①】 上記以外で、2つの集団に差があるかどうかを比較する場面を考えてみよう

--

【ワーク②】 上記の例１～３では、比較する集団にどのような違いがあるだろうか？

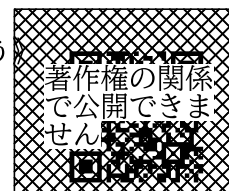
--

《仮説検定の流れ》



《表計算アプリケーションを活用して、実際に集団の比較をしてみよう》

【ワーク③】 クラスルーム/探究基礎/データサイエンス講座⑤を開き、動画を参考に平均値の検定を実施してみよう。



- 1 牛乳を習慣的に飲むことと身長が高いことに関連があるのか検証するため、幼少期から習慣的に牛乳を飲むという16歳の男子高校生8人の伸長を測定したところ、次のようになった。この母集団の平均身長が全国の16歳の男子の平均身長168.2よりも大きいといえるか、有意水準 5 %として判断せよ。

男子生徒	A	B	C	D	E	F	G	H	平均値	標本標準偏差
身長(cm)	167.1	168.9	173.4	164.5	178.1	169.8	168.3	170.1		

- 2 地域 1, 2 から切り出した同じ樹種の木材の密度(g/cm^3)をそれぞれ10ずつ調べたところ、次の通りだった。2つの地域の母平均には差があるといえるだろうか、有意水準 5% として判断せよ。

	木材の密度(g/cm^3)										平均値	標本標準偏差
地域1	0.35	0.42	0.39	0.34	0.39	0.4	0.41	0.35	0.39	0.36		
地域2	0.41	0.42	0.36	0.39	0.42	0.43	0.39	0.41	0.39	0.48		

- 3 高校生10人を対象に、あるストレッチ前後での長座体前屈の記録を測定した結果が表のとおりである。ストレッチの効果があるかどうかを、有意水準 1% として判断せよ。

生徒\記録	前(cm)	後(cm)	差
A	46.1	48.2	2.1
B	32.5	32.7	0.2
C	30.5	37.6	7.1
D	38.4	48.6	10.2
E	25.3	29.6	4.3
F	29	31.7	2.7
G	36.1	37.5	1.4
H	48.9	48.5	-0.4
I	42.3	46.9	4.6
J	39.9	43.7	3.8
平均値	***	***	3.6
標本標準偏差	***	***	3.21