

Introduction to Statistics

統計学入門

Week 13 | July 18, 2023

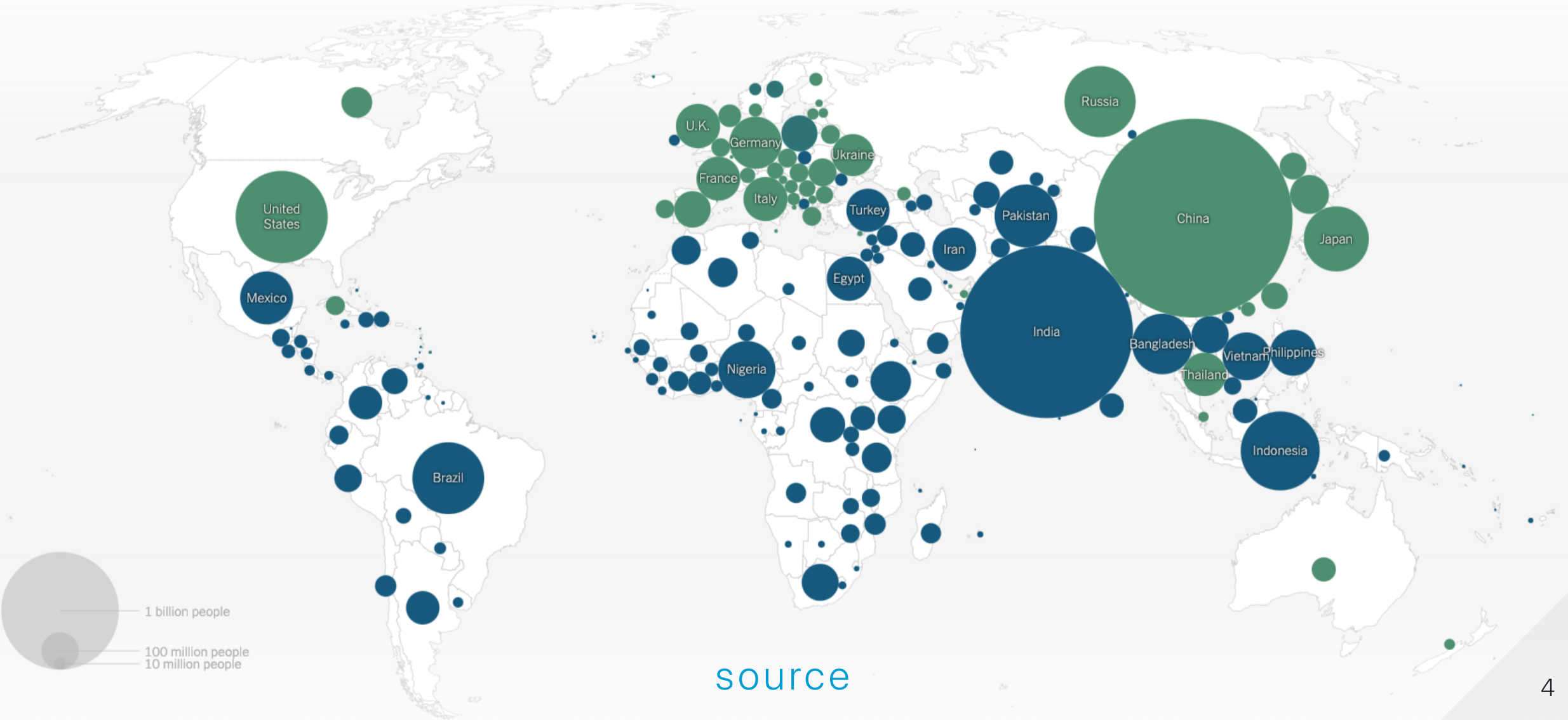
今後のスケジュール

Week	Date	Topic
13	7/18/23 (今日)	期末テスト準備パート 2
14	7/25/23	期末テスト準備パート 3
15	7/26/23(水曜11時40分~12時40分)	期末テスト

Why I want you to speak "data"

なぜ「データ」にこだわるか

1990



source

今日のまとめ課題：

平均値 と **分散** と **標準偏差** と **相関係数**

標準偏差の求め方

$$\sigma = \sqrt{\frac{1}{N} \sum_{i=1}^N (x_i - \mu)^2}$$

Diagram illustrating the components of the standard deviation formula with numbered labels:

- 7: σ (Standard Deviation)
- 5: $\frac{1}{N}$ (Reciprocal of sample size)
- 4: $\sum_{i=1}^N$ (Summation over all data points)
- 2: $(x_i - \mu)^2$ (Squared deviation from the mean)
- 1: μ (Mean)
- 3: x_i (Data point)
- 6: The entire expression inside the square root (Variance)

順番	統計	記号	式
1	平均値	μ	$\frac{1}{N} \sum x_i$
2	偏差	-	$x_i - \mu$
3	偏差平方	-	$(x_i - \mu)^2$
4	偏差平方の和	-	$\sum (x_i - \mu)^2$
5	分散	σ^2	$\frac{1}{N} \sum (x_i - \mu)^2$
6	標準偏差	σ	$\sqrt{\sigma^2}$

name	study	club	work	game	sleep	hangout	teacher	support	academic	facility	reitaku love
ね	2	0	16	20	8	2	1	1	1	1	1
わ	8	0	20	20	7	3	10	10	4	6	10
ピリー	0	2	12	10	6	4	8	8	7	7	7
柴犬	2	0	0	6	8	0	5	5	5	4	5
wwwwwwwwww	0	2	22	15	6	4	5	5	5	5	5
夕張メロン	0	6	0	4	7	0	8	4	3	3	5
アムロ	4	0	0	5	7	1	6	6	6	5	6
ああ	0	0	20	10	8	0	10	10	10	10	10
Fabrizio Romano	0	0	5	3	5	1	7	7	5	5	5
わわわ	1	1	3	3	7	2	9	8	4	7	8
お菓子泥棒	1	0	0	3	6	1	6	6	5	5	7
onigiri	0	5	0	25	7	2	4	6	5	8	6
デンジ	3	0	15	30	4	2	8	10	8	7	9
ルパン	1	0	5	5	6	0	3	3	3	6	4
ぼっぼー	0	0	20	120	5	3	8	7	8	10	9
イケメン修道士	1	2	0	45	6	0	7	7	6	7	8
chocobo	2	0	3	6	4	0	5	5	5	5	5
プクリポ	2	2	0	4	7	2	7	7	7	7	7
ルナ	5	14	0	30	8	0	5	5	7	8	9
アソバソマソ	1	2	12	40	8	0	4	5	5	6	5
たん	0	0	0	20	5	4	5	5	5	5	5
石	3	9	16	30	5	1	6	5	7	8	8
あ	30	2	20	7	6	1	5	5	5	5	9
スポンジ	3	2	16	3	4	2	8	8	7	8	8
ボブにする良い女	1	4	10	3	6	2	7	7	7	6	7

name	study	club	work	game	sleep	hangout	teacher	support	academic	facility	reitaku love
ね	2	0	16	20	8	2	1	1	1	1	1
わ	8	0	20	20	7	3	10	10	4	6	10
ビリー	0	2	12	10	6	4	8	8	7	7	7
柴犬	2	0	0	6	8	0	5	5	5	4	5
wwwwwwwww	0	2	22	15	6	4	5	5	5	5	5
夕張メロン	0	6	0	4	7	0	8	4	3	3	5
アムロ	4	0	0	5	7	1	6	6	6	5	6
ああ	0	0	20	10	8	0	10	10	10	10	10
Fabrizio Romano	0	0	5	3	5	1	7	7	5	5	5
わわわ	1	1	3	3	7	2	9	8	4	7	8
お菓子泥棒	1	0	0	3	6	1	6	6	5	5	7
onigiri	0	5	0	25	7	2	4	6	5	8	6
デンジ	3	0	15	30	4	2	8	10	8	7	9
ルパン	1	0	5	5	6	0	3	3	3	6	4
ぼっぼー	0	0	20	120	5	3	8	7	8	10	9
イケメン修道士	1	2	0	45	6	0	7	7	6	7	8
chocobo	2	0	3	6	4	0	5	5	5	5	5
ブクリポ	2	2	0	4	7	2	7	7	7	7	7
ルナ	5	14	0	30	8	0	5	5	7	8	9
アソバソマソ	1	2	12	40	8	0	4	5	5	6	5
たん	0	0	0	20	5	4	5	5	5	5	5
石	3	9	16	30	5	1	6	5	7	8	8
あ	30	2	20	7	6	1	5	5	5	5	9
スポンジ	3	2	16	3	4	2	8	8	7	8	8
ポブにする良い女	1	4	10	3	6	2	7	7	7	6	7

name	study	reitaku love
ね	2	1
わ	8	10
ピリー	0	7
柴犬	2	5
aaaaaaaaaaaa	0	5
夕張メロン	0	5
アムロ	4	6
ああ	0	10
Fabrizio Romano	0	5
わわわ	1	8
お菓子泥棒	1	7
onigiri	0	6
デンジ	3	9
ルパン	1	4
ぽっぽー	0	9
イケメン修道士	1	8
chocobo	2	5
プクリポ	2	7
ルナ	5	9
アソパソマソ	1	5
たん	0	5
石	3	8
あ	30	9
スポンジ	3	8
ボブにする良い女	1	7

name	study	平均からの 偏差	平均からの 偏差の二乗 (偏差平方)	reitaku love	平均からの 偏差	平均からの 偏差の二乗 (偏差平方)	平均からの 偏差の積
ね	2			1			
わ	8			10			
ビリー	0			7			
柴犬	2			5			
aaaaaaaaaaaa	0			5			
夕張メロン	0			5			
アムロ	4			6			
ああ	0			10			
Fabrizio Romano	0			5			
わわわ	1			8			
お菓子泥棒	1			7			
onigiri	0			6			
デンジ	3			9			
ルパン	1			4			
ぼっぼー	0			9			
イケメン修道士	1			8			
chocobo	2			5			
ブクリポ	2			7			
ルナ	5			9			
アソバソマソ	1			5			
たん	0			5			
石	3			8			
あ	30			9			
スポンジ	3			8			
ポプにする良い女	1			7			
合計							

まずは **平均値**

$$\mu = \frac{1}{N} \sum_{i=1}^N x_i$$

$$n(\text{データ数}) = 25\text{人}$$

name	study	平均からの 偏差	平均からの 偏差の二乗 (偏差平方)	reitaku love	平均からの 偏差	平均からの 偏差の二乗 (偏差平方)	平均からの 偏差の積
ね	2			1			
わ	8			10			
ビリー	0			7			
柴犬	2			5			
aaaaaaaaaaaa	0			5			
夕張メロン	0			5			
アムロ	4			6			
ああ	0			10			
Fabrizio Romano	0			5			
わわわ	1			8			
お菓子泥棒	1			7			
onigiri	0			6			
デンジ	3			9			
ルパン	1			4			
ぼっぼー	0			9			
イケメン修道士	1			8			
chocobo	2			5			
ブクリボ	2			7			
ルナ	5			9			
アソバソマソ	1			5			
たん	0			5			
石	3			8			
あ	30			9			
スポンジ	3			8			
ポプにする良い女	1			7			
合計	70			168			
平均値	2.8			6.72			

ベボシ	3		6
ボブにする良い女	1		7
合計	70		168
平均値	2.8		6.72

studyの平均 = $70/25 = 2.8$

reitakuの平均 = $168/25 = 6.72$

平均からの **偏差**

name	study	平均からの 偏差	平均からの 偏差の二乗 (偏差平方)	reitaku love	平均からの 偏差	平均からの 偏差の二乗 (偏差平方)	平均からの 偏差の積
ね	2	-0.8		1	-5.72		
わ	8	5.2		10	3.28		
ビリー	0	-2.8		7	0.28		
柴犬	2	-0.8		5	-1.72		
wwwwwwwww	0	-2.8		5	-1.72		
夕張メロン	0	-2.8		5	-1.72		
アムロ	4	1.2		6	-0.72		
ああ	0	-2.8		10	3.28		
Fabrizio Romano	0	-2.8		5	-1.72		
わわわ	1	-1.8		8	1.28		
お菓子泥棒	1	-1.8		7	0.28		
onigiri	0	-2.8		6	-0.72		
デンジ	3	0.2		9	2.28		
ルパン	1	-1.8		4	-2.72		
ぼっぼー	0	-2.8		9	2.28		
イケメン修道士	1	-1.8		8	1.28		
chocobo	2	-0.8		5	-1.72		
ブクリボ	2	-0.8		7	0.28		
ルナ	5	2.2		9	2.28		
アソバソマソ	1	-1.8		5	-1.72		
たん	0	-2.8		5	-1.72		
石	3	0.2		8	1.28		
あ	30	27.2		9	2.28		
スポンジ	3	0.2		8	1.28		
ポップにする良い女	1	-1.8		7	0.28		
合計	70			168			
平均値	2.8			6.72			

平均からの偏差の二乗
(偏差平方)

name	study	平均からの 偏差	平均からの 偏差の二乗 (偏差平方)	reitaku love	平均からの 偏差	平均からの 偏差の二乗 (偏差平方)	平均からの 偏差の積
ね	2	-0.8	0.64	1	-5.72	32.7184	
わ	8	5.2	27.04	10	3.28	10.7584	
ビリー	0	-2.8	7.84	7	0.28	0.0784	
柴犬	2	-0.8	0.64	5	-1.72	2.9584	
wwwwwwwww	0	-2.8	7.84	5	-1.72	2.9584	
夕張メロン	0	-2.8	7.84	5	-1.72	2.9584	
アムロ	4	1.2	1.44	6	-0.72	0.5184	
ああ	0	-2.8	7.84	10	3.28	10.7584	
Fabrizio Romano	0	-2.8	7.84	5	-1.72	2.9584	
わわわ	1	-1.8	3.24	8	1.28	1.6384	
お菓子泥棒	1	-1.8	3.24	7	0.28	0.0784	
onigiri	0	-2.8	7.84	6	-0.72	0.5184	
デンジ	3	0.2	0.04	9	2.28	5.1984	
ルパン	1	-1.8	3.24	4	-2.72	7.3984	
ぼっぼー	0	-2.8	7.84	9	2.28	5.1984	
イケメン修道士	1	-1.8	3.24	8	1.28	1.6384	
chocobo	2	-0.8	0.64	5	-1.72	2.9584	
ブクリボ	2	-0.8	0.64	7	0.28	0.0784	
ルナ	5	2.2	4.84	9	2.28	5.1984	
アソバソマソ	1	-1.8	3.24	5	-1.72	2.9584	
たん	0	-2.8	7.84	5	-1.72	2.9584	
石	3	0.2	0.04	8	1.28	1.6384	
あ	30	27.2	739.84	9	2.28	5.1984	
スポンジ	3	0.2	0.04	8	1.28	1.6384	
ボブにする良い女	1	-1.8	3.24	7	0.28	0.0784	
合計	70			168			
平均値	2.8			6.72			

二乗なのでマイナスはありえない！

偏差平方の合計

たん	0	-2.8	7.84	5	-1.72	2.9584	
石	3	0.2	0.04	8	1.28	1.6384	
あ	30	27.2	739.84	9	2.28	5.1984	
スポンジ	3	0.2	0.04	8	1.28	1.6384	
ポップにする良い女	1	-1.8	3.24	7	0.28	0.0784	
合計	70	0	858	168	0	111.04	
平均値	2.8			6.72			

分散

石	5	0.2	0.04	8	1.28	1.6384	
あ	30	27.2	739.84	9	2.28	5.1984	
スポンジ	3	0.2	0.04	8	1.28	1.6384	
ポップにする良い女	1	-1.8	3.24	7	0.28	0.0784	
合計	70	0	858	168	0	111.04	
平均値	2.8	分散	34.32	6.72	分散	4.44	

studyの分散 = $858/25 = 34.3$

reitakuの分散 = $111.04/25 = 4.44$

標準偏差

石	5	0.2	0.04	8	1.28	1.6384	
あ	30	27.2	739.84	9	2.28	5.1984	
スポンジ	3	0.2	0.04	8	1.28	1.6384	
ポップにする良い女	1	-1.8	3.24	7	0.28	0.0784	
合計	70	0	858	168	0	111.04	
平均値	2.8	分散	34.32	6.72	分散	4.44	
		標準偏差	5.86		標準偏差	2.11	

studyの標準偏差 = 34.32の平方根 = 5.86

reitakuの標準偏差 = 4.44の平方根 = 2.11

共分散

二つの変数の平均からの散らばりを統計的に図る

$$s_{xy} = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})$$

name	study	平均からの 偏差	平均からの 偏差の二乗 (偏差平方)	reitaku love	平均からの 偏差	平均からの 偏差の二乗 (偏差平方)	平均からの 偏差の積
ね	2	-0.8	0.64	1	-5.72	32.7184	4.576
わ	8	5.2	27.04	10	3.28	10.7584	17.056
ビリー	0	-2.8	7.84	7	0.28	0.0784	-0.784
柴犬	2	-0.8	0.64	5	-1.72	2.9584	1.376
www	0	-2.8	7.84	5	-1.72	2.9584	4.816
夕張メロン	0	-2.8	7.84	5	-1.72	2.9584	4.816
アムロ	4	1.2	1.44	6	-0.72	0.5184	-0.864
ああ	0	-2.8	7.84	10	3.28	10.7584	-9.184
Fabrizio Romano	0	-2.8	7.84	5	-1.72	2.9584	4.816
わわわ	1	-1.8	3.24	8	1.28	1.6384	-2.304
お菓子泥棒	1	-1.8	3.24	7	0.28	0.0784	-0.504
onigiri	0	-2.8	7.84	6	-0.72	0.5184	2.016
デンジ	3	0.2	0.04	9	2.28	5.1984	0.456
ルパン	1	-1.8	3.24	4	-2.72	7.3984	4.896
ぼっぼー	0	-2.8	7.84	9	2.28	5.1984	-6.384
イケメン修道士	1	-1.8	3.24	8	1.28	1.6384	-2.304
chocobo	2	-0.8	0.64	5	-1.72	2.9584	1.376
ブクリボ	2	-0.8	0.64	7	0.28	0.0784	-0.224
ルナ	5	2.2	4.84	9	2.28	5.1984	5.016
アソバソマソ	1	-1.8	3.24	5	-1.72	2.9584	3.096
たん	0	-2.8	7.84	5	-1.72	2.9584	4.816
石	3	0.2	0.04	8	1.28	1.6384	0.256
あ	30	27.2	739.84	9	2.28	5.1984	62.016
スポンジ	3	0.2	0.04	8	1.28	1.6384	0.256
ボブにする良い女	1	-1.8	3.24	7	0.28	0.0784	-0.504
合計	70	0	858	168	0	111.04	98.6
平均値	2.8	分散	34.32	6.72	2.28	5.1984	3.944

$$\text{共分散} = 98.6 \text{ (偏差の積の合計)} / 25 = 3.94$$

相関係数

$$\text{相関係数} = \frac{\text{共分散}}{\text{studyの標準偏差} * \text{reitakuの標準偏差}}$$

$$\text{相関係数} = \frac{3.94}{5.86 * 2.11} = 0.32$$

相関係数 r の値	相関
$-1 \leq r \leq -0.7$	強い負の相関
$-0.7 \leq r \leq -0.4$	負の相関
$-0.4 \leq r \leq -0.2$	弱い負の相関
$-0.2 \leq r \leq 0.2$	ほとんど相関がない
$0.2 \leq r \leq 0.4$	弱い正の相関
$0.4 \leq r \leq 0.7$	正の相関
$0.7 \leq r \leq 1$	強い正の相関

ではやってみよう

まずは紙で、残りの 9 変数の相関係数を計算する

club

work

game

sleep

hangout

teacher

support

academic

facility

ボーナスポイント **3** 点ずつ！

一枚ずつ写メでアップロード

答えは自分で Google Sheet で計算して、確認しよう

1. このシートを自分のアカウントにコピー

2. 以下の関数を打ち込んで答えを出す

action	formula
平均値	=AVERAGE(A1:A10)
分散	=VAR.P(A1:A10)
標準偏差	=STDEV.P(A1:A10)
共分散	=COVARIANCE.P(A1:A10,B1:B10)
相関係数	=CORREL(A1:A10,B1:B10)

＊ (A1:A10) はデータの幅なので当てはまる範囲に入れ替えよう

時間が余った？

プログラミング (Python) で統計やってみよう！

1. クラスサイトから `week13.ipynb` を実行
2. Google Colabで開く
3. 一緒に実行しよう！

