

- (1) 目標：データを用いて新たな科学的および社会的に有益な知見を引き出す力を向上させる。
- (2) 教材：①「分野融合～データサイエンス入門～」ワークシート (A 3 片面 各自 1 枚)  
 ②「都道府県別データセット」 (A 3 片面 各自 1 枚)  
 ③「データの分析復習」 (A 3 片面 各自 1 枚) ②③で両面印刷
- (3) 持ち物：筆記用具, AKC ファイル, タブレット PC
- (4) 事前準備：班分け(1班4人×10班を基本とする。)
- (5) 担当者：〇〇、〇〇、〇〇、〇〇  
 3 テーマでローテーション
- (6) 事前準備：エクセルデータ「都道府県別データセット」を Teams に投稿する。生徒に「前日までにエクセルデータを生徒用 TPC へダウンロードしておくこと」と伝える。
- (7) 本時の指導計画

段階	学習内容	学習活動	指導上の留意点
導入 7分	本時の内容の確認  グラフ作成実演	<ul style="list-style-type: none"> <li>・教材 2 枚を配布し、ワークシートの序文を理解する。</li> <li>・指導者の PC をプロジェクターで投影し、エクセルを用いたグラフの作成方法を理解する。</li> </ul>	<ul style="list-style-type: none"> <li>・本時は「データサイエンスの入門的な活動」を行うと伝える。</li> <li>・ほとんどの生徒は、エクセルを用いたグラフ作成は経験がない。(情報の授業では 3 学期に扱うらしい (令和 7 年度))</li> </ul>
展開 1 10分	Q 1 について	<ul style="list-style-type: none"> <li>・班ごとに席を合わせる。</li> <li>・ワークシートの会話文を読み、Q 1 に取り組む。(班で)</li> </ul>	<ul style="list-style-type: none"> <li>・紙媒体の配布資料を見ながら、どんどん意見を出させる。</li> <li>・エクセルはまだ使用せず (後で使用するので)、班での議論に時間を充てさせる。</li> </ul>
展開 2 10分	Q 2 について	<ul style="list-style-type: none"> <li>・ワークシートの Q 2 に取り組む。(班で)</li> </ul>	<ul style="list-style-type: none"> <li>・分析方法は夏課外数学「データの分析」を参考にさせる。(配布資料)</li> <li>・テーマは「科学的および社会的に有益な知見が得られる」ものが望ましい。</li> <li>・平均寿命と健康で散布図を描き、相関の有無を分析するのが扱いやすい。(平均寿命と塩分摂取量など)</li> </ul>
展開 3 20分	Q 3 について	<ul style="list-style-type: none"> <li>・ワークシートの Q 3 に取り組む。(個人で)</li> </ul>	<ul style="list-style-type: none"> <li>・エクセルの使用方法は、班員で教え合わせるか、指導者が適宜支援する。</li> </ul>
時間が余れば		<p>アイデアの案 (仮説検定は今回のデータセットでは基本的には使えない)</p> <ul style="list-style-type: none"> <li>・ある指標について、地方ごとに箱ひげ図を描いて分析する。(エクセルで描画可)</li> <li>・平均寿命と塩分摂取量で相関を調べる。また年間塩分摂取量の目安について、通説の値の妥当性について検証する。</li> <li>・米とパンの消費量から合計カロリーを計算し、多分一定になるはず?</li> </ul>	<ul style="list-style-type: none"> <li>・エクセルでグラフなどを示しながら、結果の差異があれば検討させる。</li> </ul>
まとめ 3分	本時のまとめ	<ul style="list-style-type: none"> <li>・AKC ファイルに本時で使用したプリント 2 枚を綴じる。</li> </ul>	

## 資料1：都道府県別寿命，各県基礎データ，生活時間

	都道府県	地域	平均寿命・男 (歳)	平均寿命・女 (歳)	人口 (千人)	平均年収 (万円)	面積 (km <sup>2</sup> )	平日1日当たり 睡眠時間(分)	平日1日当たり 仕事時間(分)	人口密度
1	北海道	東北	80.28	86.77	5352	417.21	83424	462	240	0.0642
2	青森	東北	78.67	85.93	1293	365.48	9646	473	252	0.1340
3	岩手	東北	79.86	86.44	1268	362.11	15275	467	255	0.0830
4	宮城	東北	80.99	87.16	2330	455.45	7282	457	252	0.3200
5	秋田	東北	79.51	86.38	1010	370.91	11638	476	236	0.0868
6	山形	東北	80.52	86.96	1113	371.48	9323	467	264	0.1194
7	福島	東北	80.12	86.4	1901	410.88	13784	460	285	0.1379
8	茨城	関東	80.28	86.33	2905	502.72	6097	446	253	0.4765
9	栃木	関東	80.1	86.24	1966	475.98	6408	453	255	0.3068
10	群馬	関東	80.61	86.84	1967	460.62	6362	452	270	0.3092
11	埼玉	関東	80.82	86.66	7289	469.97	3797	437	251	1.9197
12	千葉	関東	80.96	86.91	6236	473.18	5157	440	250	1.2092
13	東京	関東	81.07	87.26	13624	605.99	2191	443	267	6.2182
14	神奈川	関東	81.32	87.24	9145	544.28	2416	441	258	3.7852
15	新潟	中部	80.69	87.32	2286	408.79	12584	459	276	0.1817
16	富山	中部	80.61	87.42	1061	451.84	4248	455	271	0.2498
17	石川	中部	81.04	87.28	1151	435.32	4186	450	269	0.2750
18	福井	中部	81.27	87.54	782	430.28	4190	458	275	0.1866
19	山梨	中部	80.85	87.22	830	461.29	4465	454	262	0.1859
20	長野	中部	81.75	87.67	2088	449.25	13562	458	272	0.1540
21	岐阜	中部	81	86.82	2022	456.14	10621	452	257	0.1904
22	静岡	中部	80.95	87.1	3688	472.47	7777	448	252	0.4742
23	愛知	中部	81.1	86.86	7507	536.25	5173	443	261	1.4512
24	三重	関西	80.86	86.99	1808	488.31	5774	450	251	0.3131
25	滋賀	関西	81.78	87.57	1413	491.83	4017	451	258	0.3518
26	京都	関西	81.4	87.35	2605	494.17	4612	449	241	0.5648
27	大阪	関西	80.23	86.73	8833	526.53	1905	447	239	4.6367
28	兵庫	関西	80.92	87.07	5520	484.80	8401	444	248	0.6571
29	奈良	関西	81.36	87.25	1356	467.51	3691	443	230	0.3674
30	和歌山	関西	79.94	86.47	954	435.20	4725	458	244	0.2019
31	鳥取	中国	80.17	87.27	570	384.99	3507	455	264	0.1625
32	島根	中国	80.79	87.64	690	397.12	6708	465	252	0.1029
33	岡山	中国	81.03	87.67	1915	448.16	7114	450	247	0.2692
34	広島	中国	81.08	87.33	2837	464.60	8479	452	237	0.3346
35	山口	中国	80.51	86.88	1394	443.84	6112	450	235	0.2281
36	徳島	四国	80.32	86.66	750	432.51	4147	456	249	0.1809
37	香川	四国	80.85	87.21	972	433.80	1877	450	257	0.5178
38	愛媛	四国	80.16	86.82	1375	414.78	5676	454	240	0.2422
39	高知	四国	80.26	87.01	721	396.39	7104	465	242	0.1015
40	福岡	九州	80.66	87.14	5104	437.14	4986	447	249	1.0237
41	佐賀	九州	80.65	87.12	828	389.48	2441	454	268	0.3392
42	長崎	九州	80.38	86.97	1367	397.37	4132	448	247	0.3308
43	熊本	九州	81.22	87.49	1774	394.71	7409	454	268	0.2394
44	大分	九州	81.08	87.31	1160	398.33	6341	455	250	0.1829
45	宮崎	九州	80.34	87.12	1096	367.81	7735	466	242	0.1417
46	鹿児島	九州	80.02	86.78	1637	384.41	9187	460	247	0.1782
47	沖縄	九州	80.27	87.44	1439	349.88	2281	448	262	0.6309

## 資料2：医療施設，がん，喫煙，飲酒等

都道府県別データセット

	都道府県	病院数	病床数	人口10万人あたり がん死亡率(人)	年間 自殺者数	成人喫煙率 (男)(%)	成人喫煙率 (女)(%)	飲酒費用 (円)
1	北海道	575	97223	360	970	39.2	17.8	20611
2	青森	101	17916	390	282	40.3	14.3	12406
3	岩手	92	17789	368	264	38.0	10.2	22652
4	宮城	142	25488	289	412	37.3	11.9	18690
5	秋田	73	15756	412	245	38.2	10.6	19475
6	山形	68	15027	360	209	34.4	9.1	24824
7	福島	128	26137	343	374	38.9	12.1	21404
8	茨城	183	32350	305	494	35.6	11.6	18101
9	栃木	109	21704	296	363	35.5	11.2	16955
10	群馬	131	24742	306	363	35.8	10.4	15628
11	埼玉	344	62152	262	1169	33.8	13.1	20305
12	千葉	279	57109	276	982	33.6	10.7	18162
13	東京	646	127595	248	2147	31.3	11.5	29080
14	神奈川	343	74160	257	1286	30.1	9.7	20816
15	新潟	131	29169	349	461	35.0	9.1	20936
16	富山	109	17017	338	211	33.3	8.0	19262
17	石川	99	18829	312	193	34.4	9.0	21821
18	福井	71	11192	308	124	33.3	7.3	16189
19	山梨	60	11115	298	143	36.8	11.0	17079
20	長野	130	24258	310	337	31.9	8.7	24182
21	岐阜	103	20816	305	347	32.4	9.7	13928
22	静岡	182	38830	289	636	33.6	10.6	16513
23	愛知	325	67680	255	1165	33.7	9.3	13025
24	三重	103	20604	289	329	31.6	7.9	12119
25	滋賀	58	14651	260	211	32.7	7.7	14097
26	京都	173	35965	294	363	29.5	8.5	17117
27	大阪	535	108645	294	1244	33.1	12.9	14176
28	兵庫	353	65192	300	967	31.2	8.7	16588
29	奈良	75	16468	303	186	28.2	7.2	16237
30	和歌山	86	13850	348	180	32.1	7.4	10794
31	鳥取	45	8817	355	92	33.2	6.9	16181
32	島根	53	11049	368	117	32.6	6.3	22119
33	岡山	170	29378	292	263	33.4	7.8	13344
34	広島	248	40853	294	472	33.6	9.0	19947
35	山口	148	27284	345	227	31.8	8.1	17159
36	徳島	114	14936	334	117	29.9	6.1	19075
37	香川	92	15416	306	158	31.6	7.9	14957
38	愛媛	143	22779	339	291	30.8	7.7	14253
39	高知	133	18558	344	117	35.4	10.4	38910
40	福岡	465	86533	308	877	37.7	11.9	18310
41	佐賀	108	15119	335	130	39.6	10.4	17825
42	長崎	159	27124	348	238	36.9	9.4	15092
43	熊本	214	35264	316	283	35.5	8.2	23616
44	大分	158	20076	314	199	35.4	9.3	17643
45	宮崎	141	19308	327	202	35.6	9.0	21275
46	鹿児島	259	34521	324	292	33.5	8.0	20258
47	沖縄	94	18840	210	252	32.0	9.5	18268

## 資料3：犯罪・事故等

## 都道府県別データセット

	都道府県	犯罪 認知件数	検挙件数	検挙率	交通事故 発生件数	死者数 (人)	負傷者数 (人)	10万人当たり 死者数
1	北海道	25459	10777	42.3	11329	158	13489	2.94
2	青森	4060	2331	57.4	3740	53	4539	4.05
3	岩手	3458	1710	49.5	2373	73	2971	5.70
4	宮城	13755	6226	45.3	7986	71	10057	3.04
5	秋田	2460	1794	72.9	2177	54	2691	5.28
6	山形	3614	2618	72.4	6136	28	7670	2.49
7	福島	10277	4341	42.2	5802	90	7112	4.70
8	茨城	22550	8181	36.3	10455	150	13441	5.14
9	栃木	11346	4787	42.2	5484	76	6882	3.85
10	群馬	12201	6110	50.1	13574	62	17279	3.14
11	埼玉	60001	18433	30.7	27816	151	34212	2.08
12	千葉	46698	14597	31.3	18022	185	22396	2.97
13	東京	114492	37579	32.8	32412	159	37828	1.18
14	神奈川	46780	20062	42.9	27091	140	32305	1.53
15	新潟	11137	5527	49.6	4694	107	5575	4.64
16	富山	4846	2640	54.5	3466	60	4003	5.63
17	石川	4722	2146	45.4	3541	48	4150	4.16
18	福井	3197	1786	55.9	1847	51	2141	6.48
19	山梨	4123	1918	46.5	4337	35	5608	4.19
20	長野	8825	4163	47.2	8298	121	10323	5.77
21	岐阜	13232	4956	37.5	6646	90	8991	4.43
22	静岡	19659	8078	41.1	31518	137	41221	3.70
23	愛知	55080	18687	33.9	41551	212	51087	2.83
24	三重	11247	4964	44.1	6038	100	8158	5.51
25	滋賀	7967	3340	41.9	5294	53	6651	3.75
26	京都	16821	5736	34.1	8087	60	9678	2.30
27	大阪	95558	21485	22.5	37920	161	45460	1.82
28	兵庫	44233	15872	35.9	27340	152	33397	2.75
29	奈良	7764	4437	57.1	4507	47	5725	3.44
30	和歌山	4848	2891	59.6	2914	40	3528	4.15
31	鳥取	2110	1412	66.9	987	17	1243	2.96
32	島根	2631	1682	63.9	1314	28	1537	4.03
33	岡山	9509	4266	44.9	8930	79	10654	4.11
34	広島	14311	6182	43.2	9763	86	12289	3.02
35	山口	5419	2964	54.7	5401	64	6660	4.56
36	徳島	3094	1691	54.7	3579	49	4424	6.48
37	香川	5222	2646	50.7	6790	61	8434	6.25
38	愛媛	8626	4074	47.2	4497	77	5317	5.56
39	高知	4052	1652	40.8	2193	42	2447	5.77
40	福岡	36701	15114	41.2	37308	143	49917	2.80
41	佐賀	3581	2286	63.8	7783	35	10377	4.20
42	長崎	3622	2394	66.1	5652	41	7416	2.98
43	熊本	6932	3681	53.1	6151	67	7929	3.75
44	大分	3331	1885	56.6	4478	42	5862	3.60
45	宮崎	4205	2260	53.7	9015	45	10280	4.08
46	鹿児島	6704	3129	46.7	7474	65	8838	3.94
47	沖縄	6878	3919	57.0	5491	39	6661	2.72

## 資料4：食生活

## 都道府県別データセット

	都道府県	コメ消費量 (kg)	パン消費量 (g)	食塩消費量 (g)	コーヒー 消費量(g)	乳酸菌飲料 購入額(円)	うどん・そば 購入金額(円)	昆布消費量 (g)
1	北海道	87.72	38698	2195	2964	3452	5416	248
2	青森	77.87	36291	4359	2658	3402	5355	653
3	岩手	74.84	34252	2606	2705	3598	6386	669
4	宮城	61.7	35223	1900	2458	3127	5673	512
5	秋田	73.59	33123	3321	2216	3797	9034	460
6	山形	85.75	33129	4076	2411	3853	7756	558
7	福島	77.65	30730	2325	2162	4634	5614	415
8	茨城	60.2	38877	1750	2392	3712	5397	234
9	栃木	70.47	40070	1458	2356	3974	6806	323
10	群馬	75.21	42878	1768	2370	6204	7379	343
11	埼玉	71.64	45786	1763	2669	3600	6610	295
12	千葉	65.47	47458	1647	2576	3422	5242	342
13	東京	59.32	46696	1637	2300	2981	5451	296
14	神奈川	69.78	48539	1575	2551	2936	6098	372
15	新潟	85.3	42963	3041	2438	3125	5739	385
16	富山	79.18	48229	1836	2555	2703	6436	629
17	石川	74.15	47116	1886	2812	3198	6314	407
18	福井	69.37	42932	2256	2139	2226	6020	379
19	山梨	76.28	37943	1963	2070	3903	6865	261
20	長野	65.72	39576	2780	2244	4040	6854	302
21	岐阜	71.27	43730	1882	2237	4221	5295	296
22	静岡	93.43	45069	1679	1709	3275	5381	269
23	愛知	74.09	49051	1475	2467	3000	6611	276
24	三重	75.72	46677	1588	2312	2897	5688	291
25	滋賀	75.58	54749	1607	3035	2165	5847	262
26	京都	74.78	61545	1604	3427	2878	7110	353
27	大阪	75.21	55487	1470	2656	2898	6164	315
28	兵庫	64.65	58803	1385	2500	2531	6424	207
29	奈良	80.82	53798	2005	2871	3731	6429	327
30	和歌山	74.61	49157	1961	2310	4042	5721	257
31	鳥取	70.34	47452	2010	3296	2837	5398	227
32	島根	72.29	49586	1485	2645	2707	5989	404
33	岡山	61.17	56666	1484	2729	3848	5964	254
34	広島	70.45	49749	1788	3001	2874	6099	276
35	山口	70.27	44750	1825	2561	3478	5322	258
36	徳島	70.58	44979	1527	2553	2930	5987	207
37	香川	69.33	44159	1679	2601	2605	12343	248
38	愛媛	71.05	47377	1651	2678	3544	6582	206
39	高知	67.48	40496	2110	2452	3765	5470	439
40	福岡	70.82	45106	1704	2688	1823	4995	317
41	佐賀	84.05	40034	1763	2120	3123	5134	255
42	長崎	82.12	45349	2038	2108	2596	4198	377
43	熊本	72.94	37331	1928	1923	3380	4250	215
44	大分	71.82	39016	2099	2426	2194	4339	351
45	宮崎	61.23	36489	1621	1962	2710	4317	192
46	鹿児島	62.25	40567	1884	1902	4065	5084	230
47	沖縄	78.32	39174	1462	2379	3039	3230	403

## ◎分散と標準偏差

### ●偏差

$n$  個のデータ、 $x_1, x_2, x_3, \dots, x_n$  とその平均値  $\bar{x}$  に対し、各データの値と平均値との差  $x_1 - \bar{x}, x_2 - \bar{x}, x_3 - \bar{x}, \dots, x_n - \bar{x}$  を ( 偏差 ) という。ここで、偏差の平均値を求めると、

$$\begin{aligned} & \frac{1}{n} \{ (x_1 - \bar{x}) + (x_2 - \bar{x}) + (x_3 - \bar{x}) + \dots + (x_n - \bar{x}) \} \\ &= \frac{1}{n} \{ (x_1 + x_2 + \dots + x_n) - n\bar{x} \} \\ &= \frac{1}{n} (x_1 + x_2 + \dots + x_n) - \bar{x} = \bar{x} - \bar{x} = 0 \end{aligned}$$

よって、偏差の平均値は常に0となるので、偏差ではデータの散らばり具合を表すことはできない。

### ●分散と標準偏差

各データの偏差を2乗した値を考えれば、これらの値はすべて0以上であり、データの値が平均値から離れるほどその値は大きくなるので、その値の平均値をとればデータの散らばり具合を表すことができる。この値を ( 分散 ) といい、(  $s^2$  ) で表す。

分散は、計算の過程で数値を2乗するため単位が変わる。そこで、分散の正の平方根をとって単位をデータの値に合わせる。この値を ( 標準偏差 ) といい、(  $s$  ) で表す。

分散や標準偏差の値が大きいほど、データの値の散らばりが大きいといえる。

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

$$\text{標準偏差} \quad s = \sqrt{\frac{1}{n} \{ (x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \dots + (x_n - \bar{x})^2 \}}$$

<例6> 家庭菜園のミニトマトと市販のミニトマトの質量を5個ずつ測ったら以下のような結果を得た。このデータ  $x, y$  についてそれぞれの平均、分散、標準偏差を求めよ。また、その結果からどちらの方が散らばりが大きいのか考察せよ。

家庭菜園のミニトマトの質量 $x$ (g)	17	14	16	20	13
市販のミニトマトの質量 $y$ (g)	18	20	17	19	16

$$\bar{x} = \frac{1}{5} \times 80 = 16, \quad \bar{y} = \frac{1}{5} \times 90 = 18$$

$$s_x^2 = \frac{1}{5} \{ (17-16)^2 + (14-16)^2 + (16-16)^2 + (20-16)^2 + (13-16)^2 \} = \frac{1}{5} (1+4+0+16+9) = 6$$

$$\text{よって } s_x^2 = 6, \quad s_x = \sqrt{6}$$

$$s_y^2 = \frac{1}{5} \{ (18-18)^2 + (20-18)^2 + (17-18)^2 + (19-18)^2 + (16-18)^2 \} = \frac{1}{5} (0+4+1+1+4) = 2$$

$$\text{よって } s_y^2 = 2, \quad s_y = \sqrt{2}$$

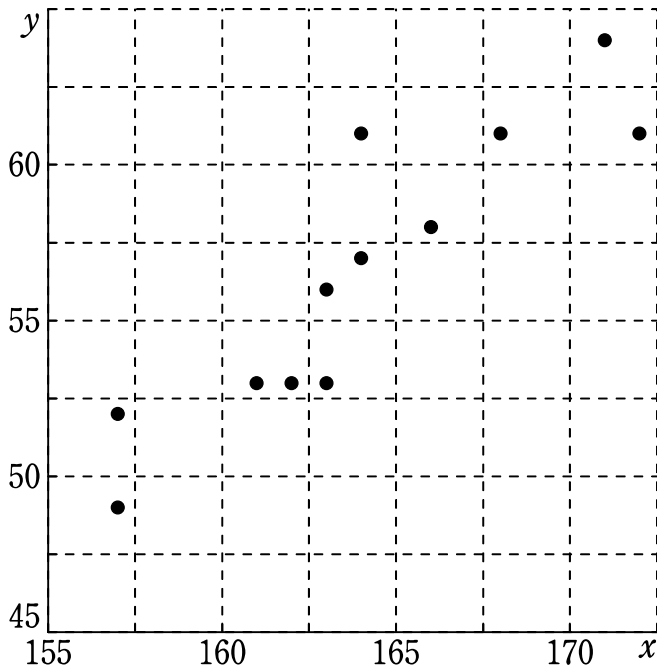
以上より  $s_x > s_y$  なので  $x$  のデータ (家庭菜園のミニトマト) の方が散らばりが大きい。

## 2. 散布図と相関係数

### ◎散布図

<例1>下の表はある高校のバスケットボール部女子12人の身長 $x(cm)$ と体重 $y(kg)$ を調べた結果である。このデータについて各 $(x, y)$ を座標とする点を以下の図に記入せよ。

番号	1	2	3	4	5	6	7	8	9	10	11	12
身長 $x$	164	171	161	162	163	172	157	157	166	164	163	168
体重 $y$	61	64	53	53	53	61	52	49	58	57	56	61



このような図を  
( 散 布 図 )  
という。

この散布図においては、対応する $x$ と $y$ の値は一方が大きいほど他方も大きい傾向にある。このとき、2つの変数 $x$ と $y$ の間に（ 正の相関関係がある ）という。また、対応する $x$ と $y$ の値は一方が大きいほど他方が小さい傾向にあるとき、2つの変数 $x$ と $y$ の間に（ 負の相関関係がある ）という。どちらの傾向も見られないときは（ 相関関係がない ）という。散布図においては、点の分布が直線状に近づくほど（ 相関関係が強い ）といい、直線状ではなく広く散らばるほど（ 相関関係が弱い ）という。

### ◎相関係数

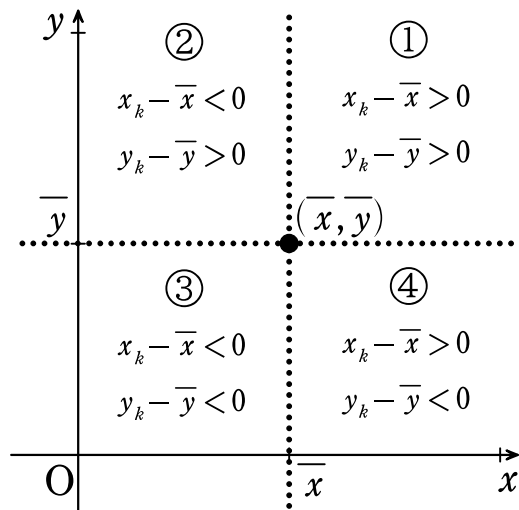
対応する2つの量の関係は散布図によってある程度の判断はできるが、さらにその相関の程度を数値で表すことを考えてみよう。

対応する2つの変数の組を $(x_1, y_1)$ ,  $(x_2, y_2)$ ,  $(x_3, y_3)$ , ...,  $(x_n, y_n)$ とし、 $x$ と $y$ の平均値を $\bar{x}$ ,  $\bar{y}$ とする。これら $n$ 個の点は散布図にかいたとき、点 $(\bar{x}, \bar{y})$ の周りに集まっている。

図のように平面を点 $(\bar{x}, \bar{y})$ を通り座標軸に平行な直線で4つの部分①~④の部分に分ける。

$x$  と  $y$  の間に

正の相関があるとき、点は ( ① ) と ( ③ ) に  
負の相関があるとき、点は ( ② ) と ( ④ ) に  
に多く集まる傾向がある。



点  $(x_k, y_k)$  が

①または③にあるとき、

$$(x_k - \bar{x})(y_k - \bar{y}) > 0$$

②または④にあるとき、

$$(x_k - \bar{x})(y_k - \bar{y}) < 0$$

であるから、 $(x_k - \bar{x})(y_k - \bar{y})$ の平均値 は

正の相関があれば( 正の値 )、負の相関があれば( 負の値 )、相関がなければ ( 0 ) に  
近い値となる。この値を  $x, y$  の( 共分散 )といい、(  $s_{xy}$  )で表す。

すなわち

$$s_{xy} = \frac{1}{n} \{ (x_1 - \bar{x})(y_1 - \bar{y}) + (x_2 - \bar{x})(y_2 - \bar{y}) + \dots + (x_n - \bar{x})(y_n - \bar{y}) \}$$

共分散は変数  $x, y$  の散らばり具合に影響を受けるから、共分散  $s_{xy}$  を  $x, y$  の標準偏差  $s_x$  と  $s_y$  の  
積  $s_x s_y$  で割った値を考え、この値を  $x, y$  の( 相関係数 )といい、記号(  $r$  )で表す。

$$\text{相関係数} \quad r = \frac{s_{xy}}{s_x \cdot s_y}$$

※共分散  $s_{xy}$  はデータ  $x_1 y_1, x_2 y_2, \dots, x_n y_n$  の平均値  $\overline{xy}$  を用いて、

$$s_{xy} = \overline{xy} - \bar{x} \cdot \bar{y} \quad \text{で求めることもできる。}$$

相関係数  $r$  の値については

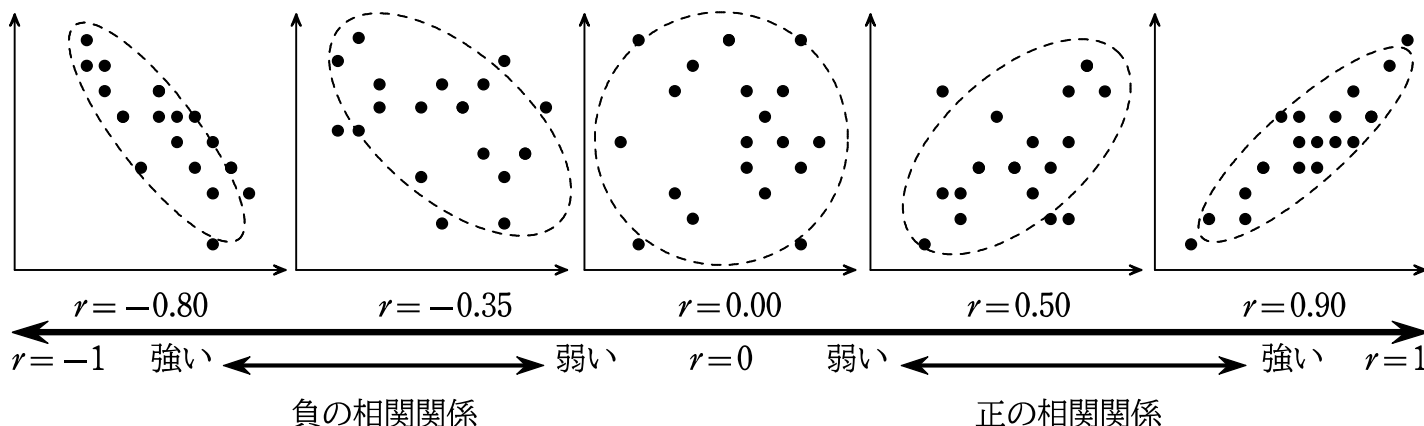
$$-1 \leq r \leq 1$$

が成り立ち、

正の相関関係が強いほど  $r$  の値は ( 1 に 近づき )、

負の相関関係が強いほど  $r$  の値は ( -1 に 近づく )。

また、 $r$  の値が ( 0 に 近いほど ) 相関は弱い。



### 3. 仮説検定の考え方

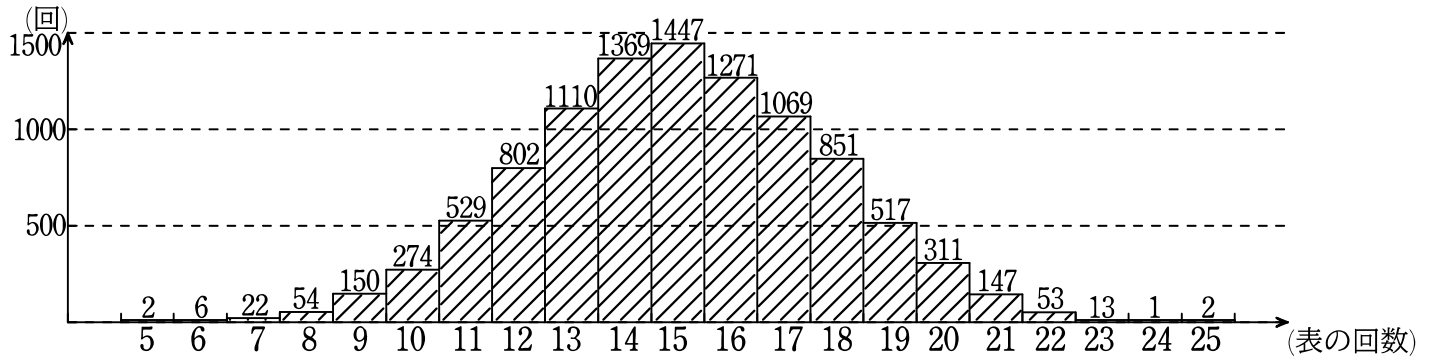
#### ◎仮説検定

相撲の勝敗予想がよく当たるといふ猫に、30の取り組みの勝敗を予想させたところ19の取り組みにおいて勝敗が的中した。この結果からこの猫の予想はよく当たると判断してよいだろうか？

「この猫の予想はよく当たる」といふ仮説Aに反する「この猫は当てずっぽうで予想している」といふ仮説Bをたて、この仮説Bのもとである事象が起りうる確率にもとづいて「この猫の予想はよく当たる」と判断できるかどうかを検討してみる。このような方法を(仮説検定)という。

仮説検定では、基準となる確率を5%とし(10%や1%に取る場合もあるが)基準よりも小さい確率で起こる事象は、ほとんど起こりえない事象であるとする。

当てずっぽうの予想が的中するかどうかを、コイン投げに置き換えて表が出れば予想が的中し、裏が出れば予想が外れたと考えることにする。以下はコンピューターを用いてコインを30回投げの実験を10000回繰り返したときに表が出た回数を表すグラフである。



このグラフを用いて考えてみよう。表が19回以上出た回数は

$517 + 311 + 147 + 53 + 13 + 1 + 2 = (1044)$  回であるから  
表が19回以上出る確率はおおよそ(10.4%)といえる。この結果から、当てずっぽうで予想して30の取り組み中19の取り組みの結果が的中する確率はおおよそ(10.4%)であり、これは5%よりも(高い)から、「当てずっぽうで予想している」といふ仮説Bのもとで十分起りえる事象と考えると、仮説を(否定するのは難しい)。

従って、この場合、「この猫の予想はよく当たる」と判断することは(できない)。

【注】この場合、仮説Bが正しいと判断できるわけではないことに注意。

「仮説A,仮説Bのどちらが正しいとも言えない」といふのが得られる結論となる。

一方、30の取り組み中、21の取り組みの結果が的中する場合はどうであろうか。

表が出た回数が21回以上になる場合は

$147 + 53 + 13 + 1 + 2 = (216)$  回であるから

当てずっぽうで予想して30の取り組み中21の取り組みの結果が的中する確率はおおよそ(2.2%)であり、これは5%よりも(低い)から、「当てずっぽうで予想している」といふ仮説Bのもとではほとんど起こりえない事象と考えると、仮説は(否定される)。  
従って、この場合、「この猫の予想はよく当たる」と判断することが(できる)。

### 分野融合～データサイエンス入門～

データサイエンスとは、データを用いて新たな科学のおよび社会的に有益な知見を引き出そうとするアプローチのことを言います。本時は、データサイエンスを実際に体験します。都道府県別データセットの中から「社会的に有益な知見」を引き出してみましよう。

**会話文** (【配布資料】都道府県別データセット資料1を見たAさんとBさんの会話)

A: 男性の平均寿命は、だいたい80歳ぐらいなんだね。

B: 平均寿命のトップは、男性で滋賀県、女性で岡山県だとニュースで聞いたことがあるな。

A: 表を見ると、どこがトップかなんてほとんど差がないから、わざわざ報道するほどのことかな？

80歳の中の1,2歳の差なんて、 $1 \div 80 = 0.125 (1.25\%)$ しかないよ。

B: でもグラフ(ヒストグラム)にすると、上位2県は他とは差があるように見えるな。

A: そういえば、夏の課外で「仮説検定」ってやったよね？

有意な差が・・・ってやつ。(【配布資料】都道府県別データセット裏面参照)

B: それって今回の場合は使えるの??? (※ちなみに使えません)

Q1: **会話文**のように「平均寿命の上位は他県と有意な差があるのか」のような、調べると何か有益な発見ができそうな具体的なテーマを、【配布資料】都道府県別データセットから5個以上探しましょう。  
(班で話し合い)

参考: 多いほどいいので、どんどんテーマを出しましょう。(理想は10個以上) 別々のテーマを組み合わせ、新しいテーマができることもあります。また、実際にエクセル等で具体的に分析せずに、ざっと見た数値の印象で答えていきましょう。

- ・
- ・
- ・
- ・
- ・

Q2：分析をするテーマについて、優先度の高いもの2つをQ1より選び、エクセルを用いた分析方法についても考えましょう。

(班で話し合い)

参考：分析方法はデータの分析（夏の課外）で習った内容も使えます。散布図をエクセルで描いて、相関の有無を見るのがやりやすいでしょう。また、追加のデータは用いてもいいですが、きっと探すのに時間がかかるので注意してください。(非推奨)

1.

テーマ

分析方法

2.

テーマ

分析方法

Q3：Q2を参考に、実際にエクセルで分析してみましょう。分析したテーマを1つ選び、その分析結果について下記にまとめましょう。

(個人で分析に取り組む)

テーマ

分析方法と結果