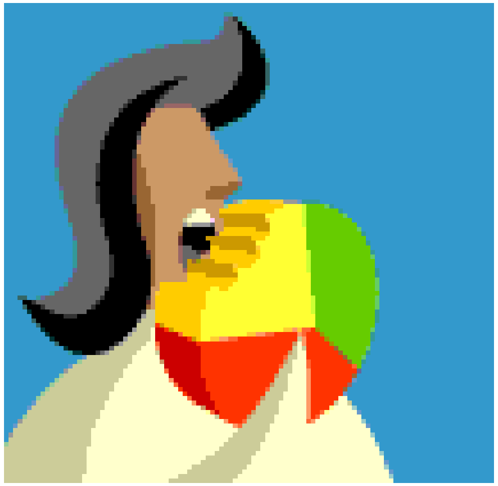
統計をどう使うか？



電子情報工学科
伊庭　斉志

統計的検定 [付録B,p.179]

- t-検定：　２つの正規母集団の平均値の差が有意かどうかを検定する

- f-検定：２つの正規母集団の分散値が等しいかどうかを検定する

t-検定の手続き

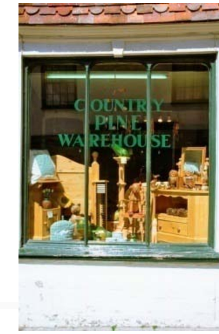
- 帰無仮説･対立仮説を立てる
  - 平均が異なるか？
- 分散の等質性を検討(f-検定)
- 条件の平均値の差についてt値を算出
- そのt値が有意であるかどうかを調べて、仮説を選択
  - t分布(t統計量の分布)のなかでどの位置にあるか
  - t分布表から有意と言えるかを判断

帰無・対立仮説

- 平均値を差を検定するt検定の帰無・対立仮説は以下のようになる
  - 帰無仮説：2条件の平均には差がない(2条件の平均値の差が0である)
  - 対立仮説：2条件の平均には差がある(2条件の平均値の差が0ではない)

待ち行列の例

VS



- フランチャイズ店のリストラを考えている
- A店とB店の2か所について経営方針（店長の質）を査定したい

A店とB店の呼損率を調べた

ランダムな日を選んでに10回計測

↓

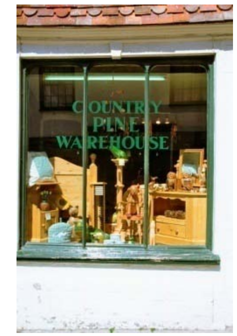
t検定

待ち行列の例

A店の呼損率
24.136
25.100
25.483
24.666
24.873
25.883
25.624
25.561
25.199
25.038

平均:25.236
分散:0.2432

B店の呼損率
23.394
21.620
23.772
23.228
23.747
23.957
23.387
22.939
24.271
23.235



平均:23.355
分散:0.5724

t値の計算法

- ２つの条件について対応のないデータの場合（教科書p.178）

$$t = \frac{\overline{X}_1 - \overline{X}_2}{\sqrt{\hat{\sigma}^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \qquad \hat{\sigma}^2 = \frac{(n_1 - 1)\hat{\sigma}_1^2 + (n_2 - 1)\hat{\sigma}_2^2}{(n_1 - 1) + (n_2 - 1)}$$

計算の結果

- t値を求める

$$t=\frac{\overline{X}_1-\overline{X}_2}{\sqrt{\hat{\sigma}^2\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}}=\frac{25.236-23.355}{\sqrt{\hat{\sigma}^2\left(\frac{1}{10}+\frac{1}{10}\right)}}=6.771$$

$$\hat{\sigma}^2=\frac{(n_1-1)\hat{\sigma}_1^2+(n_2-1)\hat{\sigma}_2^2}{(n_1-1)+(n_2-1)}=\frac{9\times 0.2432+9\times 0.5274}{9+9}$$

- 自由度を求める
  →標本数は10と10なので自由度は10+10-2=18

検定の結果をt分布表で調べる

自由度	α				自由度	α			
f	0.1	0.05	0.02	0.01	f	0.1	0.05	0.02	0.01
1	6.314	12.706	31.821	63.657	52	1.657	2.007	2.400	2.674
2	2.920	4.303	6.965	9.925	54	1.674	2.005	2.397	2.670
3	2.353	3.182	4.541	5.841	56	1.673	2.003	2.395	2.667
4	2.132	2.776	3.747	4.604	58	1.672	2.002	2.392	2.663
5	2.015	2.571	3.365	4.032	60	1.671	2.000	2.390	2.660
6	1.943	2.447	3.143	3.707	62	1.670	1.999	2.388	2.657
7	1.895	2.365	2.998	3.499	64	1.689	1.998	2.386	2.655
8	1.860	2.306	2.896	3.355	66	1.668	1.997	2.384	2.652
9	1.833	2.262	2.821	3.250	68	1.668	1.995	2.382	2.650
10	1.812	2.228	2.764	3.169	70	1.667	1.994	2.381	2.648
11	1.796	2.201	2.718	3.106	72	1.666	1.993	2.379	2.646
12	1.782	2.179	2.681	3.055	74	1.666	1.993	2.378	2.644
13	1.771	2.160	2.650	3.012	75	1.665	1.992	2.376	2.642
14	1.761	2.145	2.624	2.977	78	1.665	1.991	2.375	2.640
15	1.753	2.131	2.602	2.947	80	1.664	1.990	2.374	2.639
16	1.746	2.120	2.583	2.921	82	1.664	1.989	2.373	2.637
17	1.740	2.110	2.567	2.898	84	1.663	1.989	2.372	2.636
18	1.734	2.101	2.552	2.878	80	1.663	1.988	2.370	2.634
19	1.729	2.093	2.539	2.851	88	1.662	1.987	2.369	2.633
20	1.725	2.086	2.528	2.845	90	1.662	1.987	2.368	2.632
21	1.721	2.080	2.518	2.831	92	1.662	1.986	2.368	2.630
22	1.717	2.074	2.508	2.819	94	1.681	1.986	2.367	2.629

検定の結果をt分布表で調べる

自由度 f	α 0.1	0.05	0.02	0.01
1	6.314	12.706	31.821	63.657
2	2.920	4.303	6.965	9.925
3	2.353	3.182	4.541	5.841
4	2.132	2.776	3.747	4.604
5	2.015	2.571	3.365	4.032
6	1.943	2.447	3.143	3.707
7	1.895	2.365	2.998	3.499
8	1.860	2.306	2.896	3.355
9	1.833	2.262	2.821	3.250
10	1.812	2.228	2.764	3.169
11	1.796	2.201	2.718	3.106
12	1.782	2.179	2.681	3.055
13	1.771	2.160	2.650	3.012
14	1.761	2.145	2.624	2.977
15	1.753	2.131	2.602	2.947
16	1.746	2.120	2.583	2.921
17	1.740	2.110	2.567	2.898
18	1.734	2.101	2.552	2.878
19	1.729	2.093	2.539	2.851
20	1.725	2.086	2.528	2.845
21	1.721	2.080	2.518	2.831
22	1.717	2.074	2.508	2.819

有意水準: 0.05 = 5%

自由度＝10＋10－2＝18

$$t=\frac{\bar{X}_1-\bar{X}_2}{\sqrt{\hat{\sigma}^2\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}}=\frac{25.236-23.355}{\sqrt{\hat{\sigma}^2\left(\frac{1}{10}+\frac{1}{10}\right)}}=6.771$$

6.771>2.101なので
有意水準5%で差は有意である。
∴ 差が認められる

∴A店の呼損率は有意に多いので、リストラすべき。

t検定の適用条件

- 正規分布
- 処理効果の加法性
- 等分散性

分散が等しいかどうか

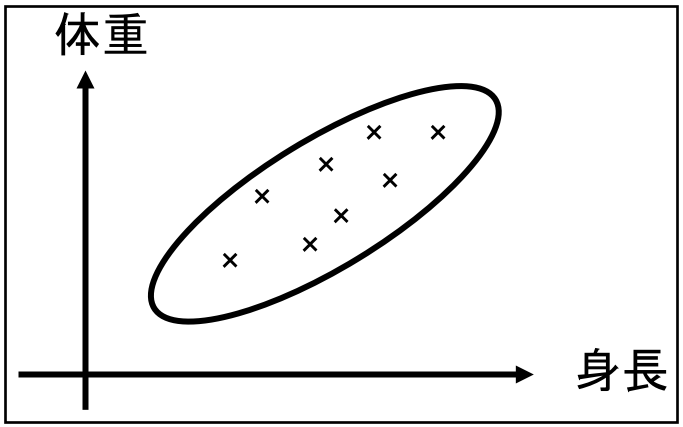
- 2群の分散は母集団において等しいとは限らない。平均値の差をt検定によって確認するのと同じように，分散が等しいか否かを確認する
- F検定を行って確認する（教科書 p.180)
- 多くの場合，分散を等しいと見なしても，等しくないと見なしてもt検定の結果にはあまり影響がない
- 対応のあるデータの場合には分散が等しいか否かは問題にならない

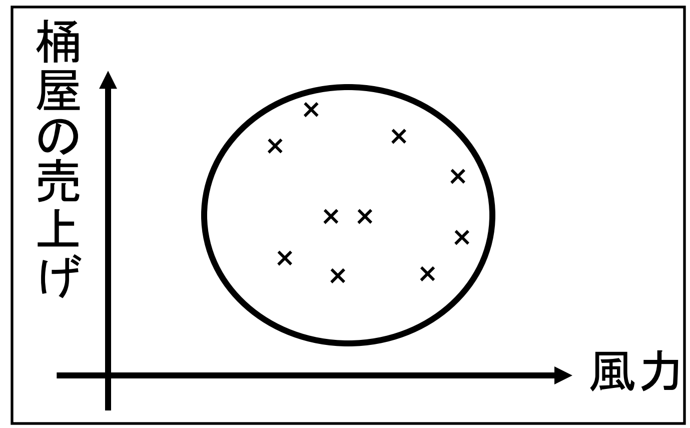
統計は嘘をつく

- 偽の相関
- 第3の要因
- 因果関係と相関関係

相関係数

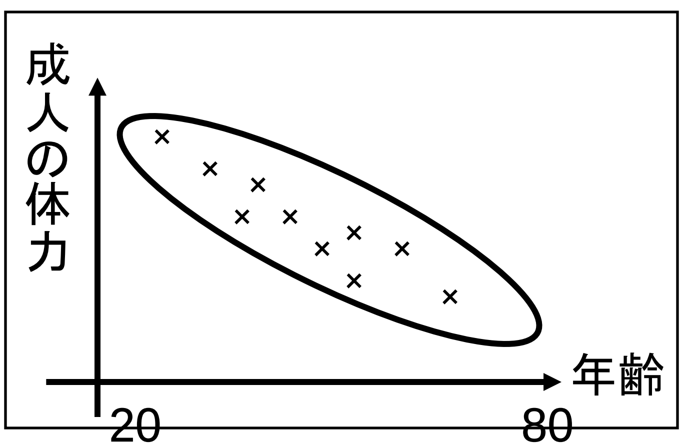
正の相関関係

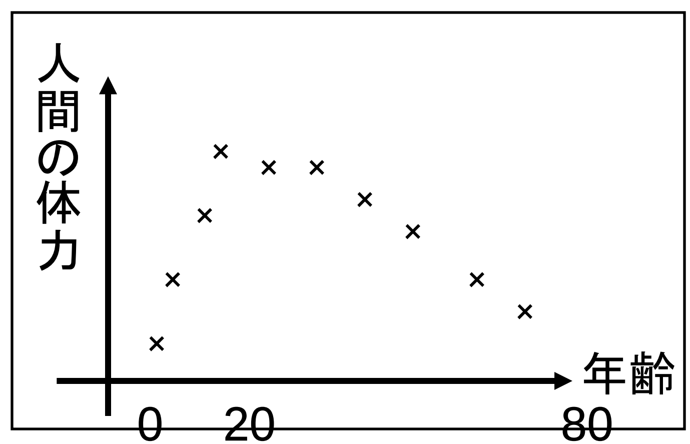




無相関

負の相関関係

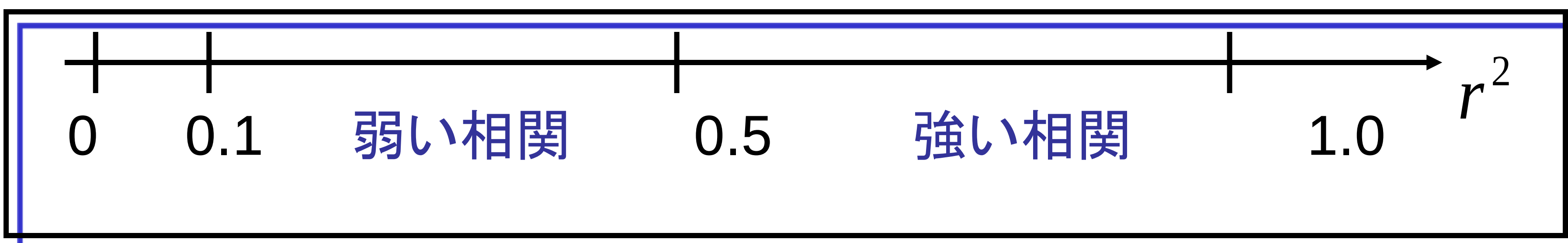




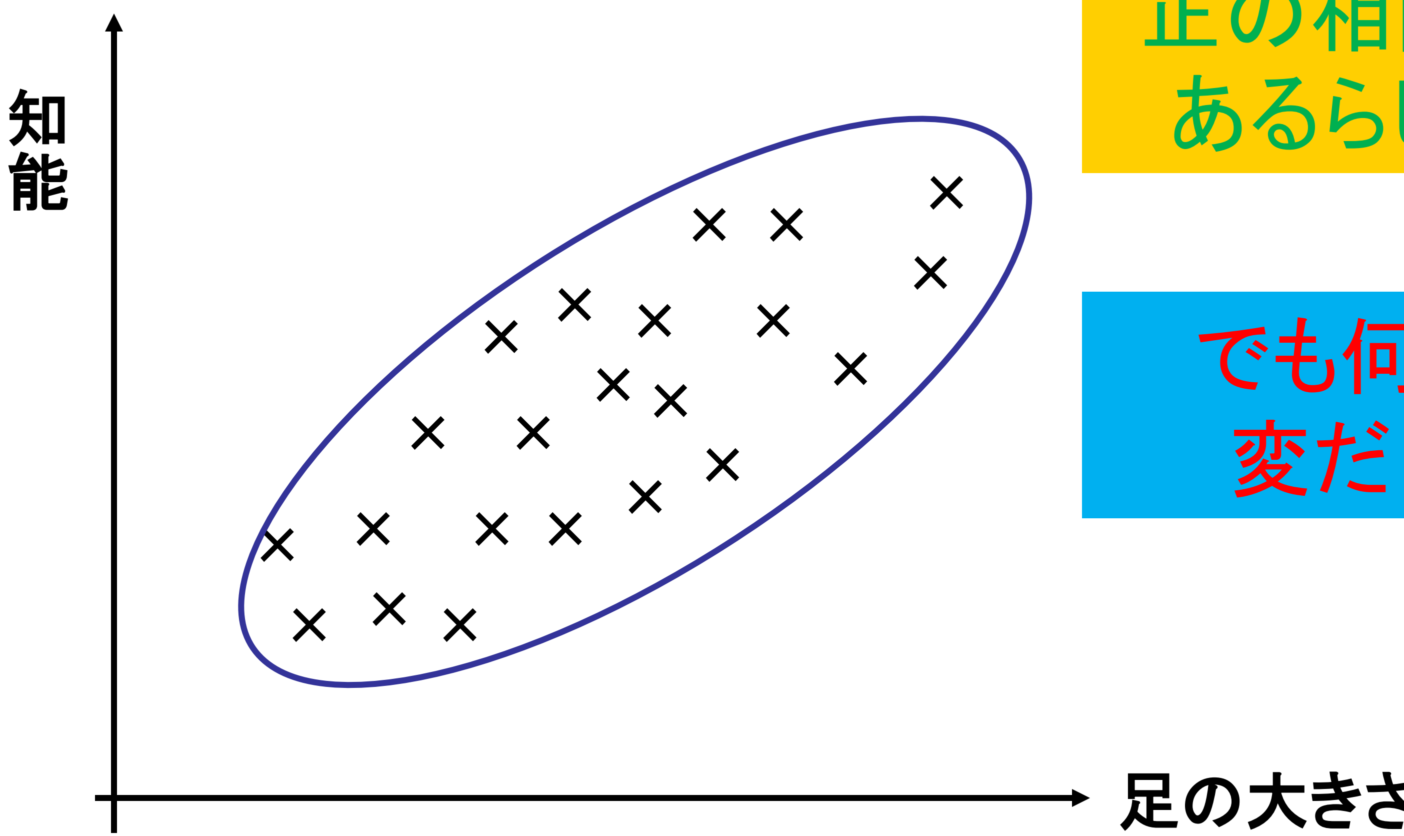
非線型な相関関係

相関係数r の見方（目安）

- $0.7<r<1.0$ ⇒ 強い正の相関
- $0.3<r<0.7$ ⇒ 弱い正の相関
- $-0.3<r<0.3$ ⇒ 無相関
- $-0.7<r<-0.3$ ⇒ 弱い負の相関
- $-1.0<r<-0.7$ ⇒ 強い負の相関



例：足の大きさと知能の相関？

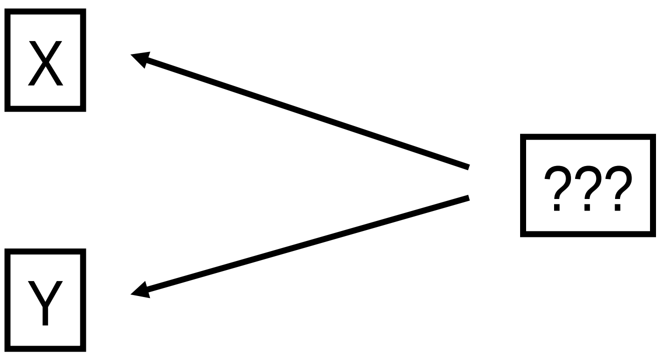


正の相関があるらしい

でも何か変だ？

擬相関とは

- 第３変数によって引き起こされる相関
- YとXが直接関係しているわけではなく、背後に存在する因子のため見かけ上の相関が生じている現象のこと



相関関係と因果関係

- 相関関係は必ずしも因果関係を意味しない
- 因果関係ある $\Rightarrow$ 相関関係ある
- 相関関係ある $\not\Rightarrow$ 因果関係ある

相関関係と因果関係

- ＡとＢに相関関係があることは、次の４つのいずれかを意味する

  - A（原因）→B（結果）
  - B（原因）→A（結果）

  擬相関

  - X（原因）→A（結果） および X（原因）→B（結果）
  - ただの偶然

統計にだまされるな

- 統計でうそをつく方法
- 相関関係を因果関係として言いくるめる
  - ヨーグルトと長寿
  - ゲーム脳の恐怖
  - 朝食をとると頭にいい？
  - 。。。

もっと知りたい人へ

- 『人間の測りまちがい　差別の科学史』by スティーブン・J・グールド
- 科学的な誤謬と差別の歴史
  - 過去の頭蓋計測学や進化論
  - 犯罪心理学
  - 知能テスト（IQ恒常説）

偉大な科学者ですら間違える!!