

データを科学的に見ること、考えること

—— 昨年の医療情報学実習発表会でのアンケートから

島根大学医学部医療情報学講座
安田晃

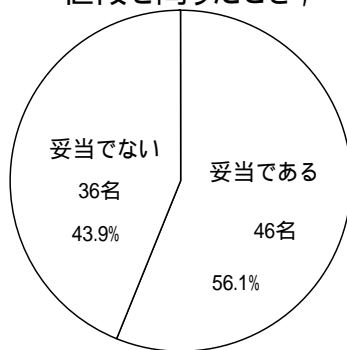
2004年4月14日(水)

科学的とは・・・

事実そのものに裏づけられ、論理的認識によって媒介されているさま。原理的に体系づけられているさま。学問的。(広辞苑第4版 CD-ROM版)

今、何らかのデータがあるとき、上の定義から何が言えるだろう。

例：この学年82名に学食のメニューと値段を問うたとき、



であったとする。

この結果から、学食のメニューと値段が妥当であると結論付けていいか？

- いいだろう。何となれば過半数を超えている。
- 確かに半数以上だが、超え方が少し微妙だ。例えば56%だと、僅かの超過にしか見えない。
- 確かに46名は半分以上だが、半分以上と確約できる数字は、82名という全体の数字に依存するだろう。

では、どこまでいけば妥当と認められるのだろう？

- 人によって過半数の概念が異なるから、一意に定まらない。
- どこかに経験則があるだろう。
- 統計学的な検定があるだろう。

正解は3番目の意見。

平均値、標準偏差、偏差値など標本を代表する値、2変数の関係、拡張して P 変数の関係、類似性、背反性、検定など、統計学の言葉で話せば、分かりやすい。

——— この問題はこのようにして解ける(あくまで一例)。

妥当が半分を超えているか？ ()

比率 p で起こる事象が, n 回の試行で k 回起こったときの比率 p_k と有意に異なるか, という検定は,

$$v_1 = 2(n - k + 1), v_2 = 2k$$

とおいて,

$$F_0 = \frac{v_2(1-p)}{v_1 p}$$

なる統計量が,

$$F_0 \geq F_{v_2}^{v_1}(\alpha)$$

ならば,

p と p_k の差は有意である. ここで,

$F_{v_2}^{v_1}(\alpha)$ は, 自由度 v_1, v_2 , 有意水準 α の F 値

妥当が半分を超えているか？ ()

上述の問題では, $p = \frac{1}{2}$ とする.

$$F_0 = \frac{v_2(1-p)}{v_1 p} = \frac{\frac{1}{2} v_2}{\frac{1}{2} v_1} = \frac{k}{n - k + 1} \geq F_{v_2}^{v_1}(\alpha)$$

と書ける.

ここで有意水準 α を 0.05 とする. そうすれば, 以下の表を得る.

妥当が半分を超えているか？ ()

総数	妥当者数	F_0	v_1	v_2	$F_{v_2}^{v_1}(0.05)$
82	42	1.024	82	84	1.437
	43	1.075	80	86	1.436
	44	1.128	78	88	1.436
	45	1.184	76	90	1.435
	46	1.243	74	92	1.435
	47	1.306	72	94	1.435
	48	1.371	70	96	1.436
	49	1.441	68	98	1.437
	50	1.515	66	100	1.438
	51	1.594	64	102	1.439
	52	1.677	62	104	1.441
	53	1.767	60	106	1.443
	54	1.862	58	108	1.446

ここで, F_0 が上回った. 49人以上いると妥当が多いと言っていいだろう.

妥当が半分を超えているか？ ()

故に, 妥当であると答えたものは, 多いとは言えないことが, 有意水準 0.05 で統計学的に分かった.

医療, 医学 (に限らないが) は科学的に, 合理的にデータを見つめてほしい. そして, 考察してほしい.

このように前置きして, 本題へ進む.

本題

昨年度の医療情報実習発表会で, あなたたちは各グループをどのように評価したか.

医療情報学実習発表会では,

- 21グループが発表順を決め,
- 発表10分, 質疑討論5分で発表を行い,
- 発表終了後直ちに, そのグループの採点を, 5点満点の整数で行った. 自身のグループは評価していない.

21グループは、

- 診療録について考えるA, B (診療録)
- クリニカルパスA, B (クリパス)
- 遠隔地医療A, B (遠隔地)
- 医療分野における個人情報保護A, B (情報保護)
- 老人、障害者外来受診に必要な設備 (外来受診)
- チーム医療A, B (チーム医療)
- 診療報酬の包括支払制度について (包括支払)
- 高齢化社会と年金保険制度A, B (年金保険)
- 病院の経営分析(経営分析)
- データ解析によって見た現在の福祉、介護A, B (福祉介護)
- 医療に対する考え方をアンケートから探る(医学科編)A, B (質問紙医)
- 医療に対する考え方をアンケートから探る(看護学科編)A, B (質問紙看)

2003年度医療情報学実習自由研究発表会採点表

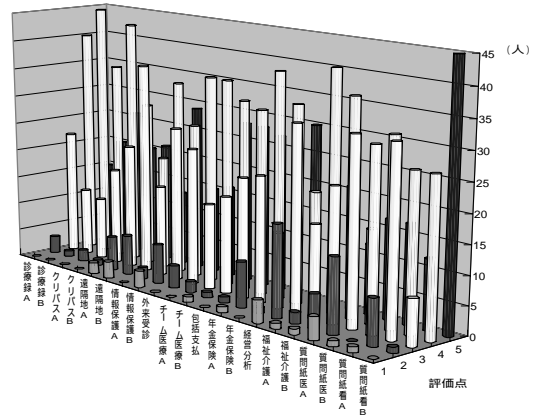
発表内容は適切でしたか (内容適切)	点
よく勉強して発表していましたか (勉強発表)	点
プレゼンテーションの方法は (プレゼン)	点
質問への対応は適切でしたか (対応適切)	点
今後に生かせる内容でしたか (将来性)	点

データを眺めてみる(1)

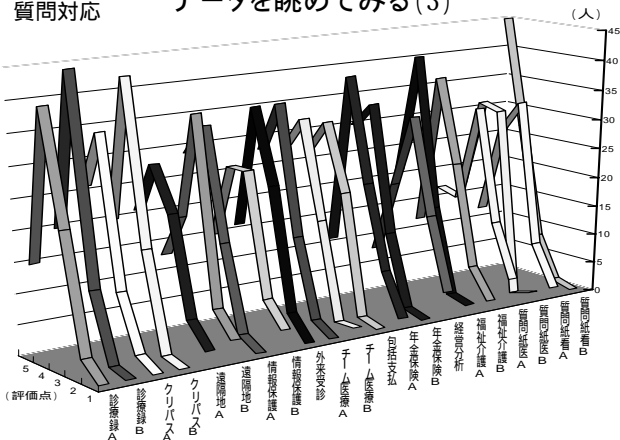
診療録A

	内容適切	勉強発表	プレゼン	対応適切	将来性
0点	0	0	0	0	0
1点	0	0	0	0	0
2点	2	1	0	3	3
3点	10	12	28	22	16
4点	34	34	36	40	33
5点	35	34	17	15	29
合計	81	81	81	80	18

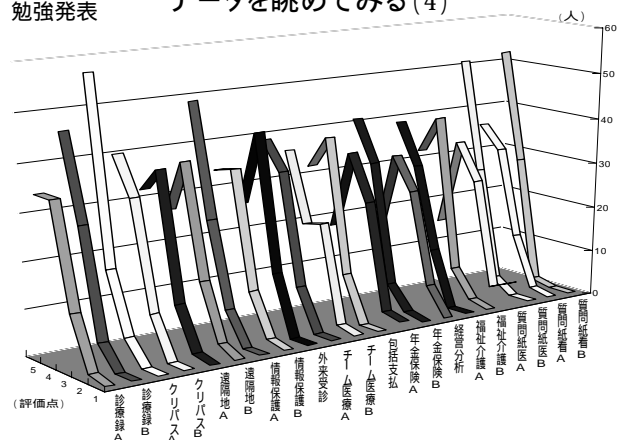
質問対応 データを眺めてみる(2)



質問対応 データを眺めてみる(3)



勉強発表 データを眺めてみる(4)



データを眺めてみる(11)

チュータと学生は同一な評価をしているのだろうか。

質問対応の総得点をチュータ、学生ともに順位で表わすと、

	診A	診B	クA	クB	遠A	遠B	情A	情B	外	チA	チB	包A	包B	年A	年B	経A	経B	福A	福B	医A	医B	看A	看B
T	16	10	4	5	16	18	8	16	11	4	10	8	13	2	21	8	13	20	18	20	1		
S	12	9	12	7	14	20	3	18	13	3	6	4	15	5	21	8	16	19	12	18	1		

このデータを計算して、順位相関係数 r_s は

$$r_s = 0.8792$$

すべての質問項目について行えば、

データを眺めてみる(11)

チュータと学生は同一な評価をしているのだろうか。

質問項目	順位相関係数
内容適切	0.7091
勉強発表	0.8945
プレゼン	0.3058
質問対応	0.8792
将来性	0.7948

プレゼンを除いて有意水準0.05以下であった。

データを解析する(1)

今までの解析結果は単変量の断片的なものだった。
これからは少し複雑な解析結果を紹介し、医療情報学自由研究発表会での行動を考えてゆきたい。

データを解析する(2)

内容適切な集計表は、

	2点	3点	4点	5点
診療録A	2	10	34	35
診療録B	0	2	27	50
クリバスA	0	5	24	51
クリバスB	0	3	42	36
遠隔地A	0	11	39	30
遠隔地B	1	13	31	36
情報保護A	0	10	32	38
情報保護B	1	15	36	28
外来受診	0	10	42	29
チーム医療A	0	5	35	38
チーム医療B	0	8	35	37

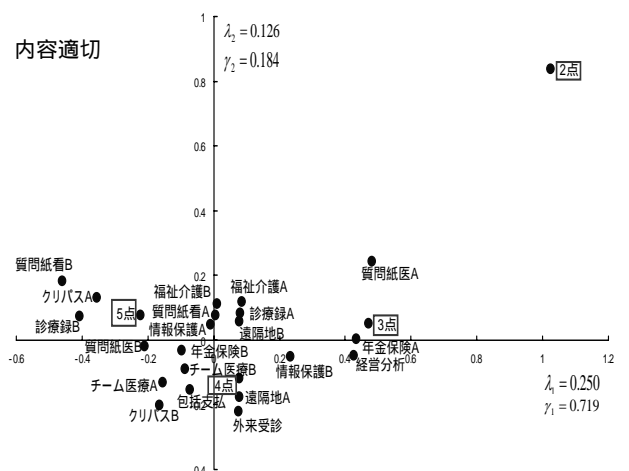
	2点	3点	4点	5点
包括支払	0	7	39	35
年金保険A	2	19	37	23
年金保険B	0	9	32	39
経営分析	1	21	35	23
福祉介護A	0	11	32	36
福祉介護B	1	12	28	36
質問紙医A	5	16	37	25
質問紙医B	0	6	32	43
質問紙看A	1	11	30	38
質問紙看B	0	3	22	56

数字は人

データを解析する(3)

このデータから、評価点である列の要素とグループである行の要素を対応させたい。

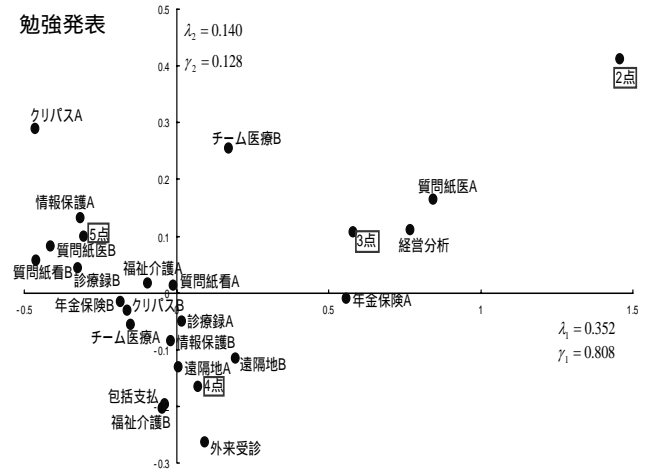
そこで対応分析を行って、第2固有値までの固有ベクトルをプロットすれば、



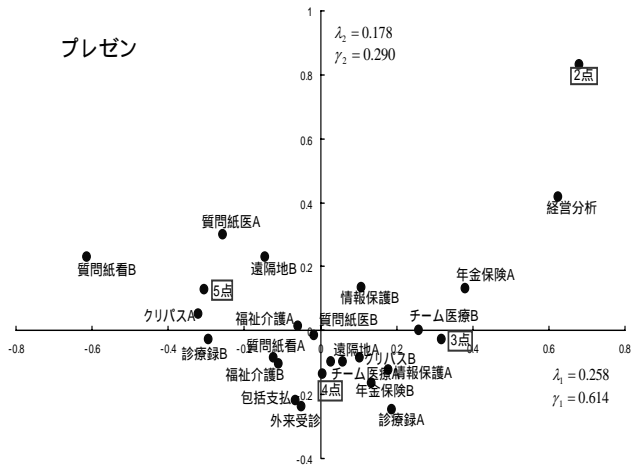
データを解析する(5)

この他の結果も紹介する．自身が所属したグループ
がどの位置にいるか探してほしい．

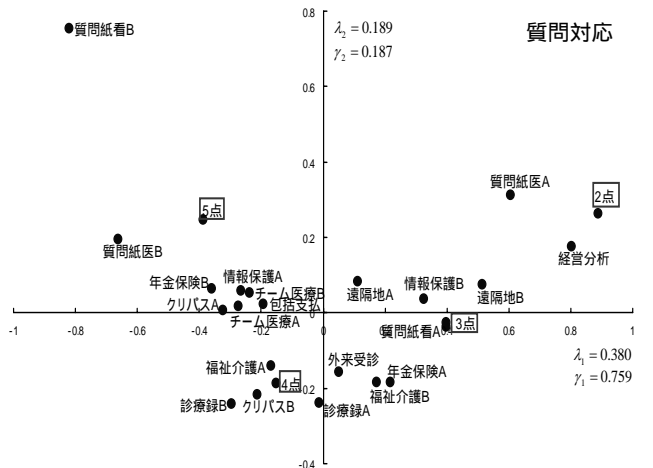
勉強発表



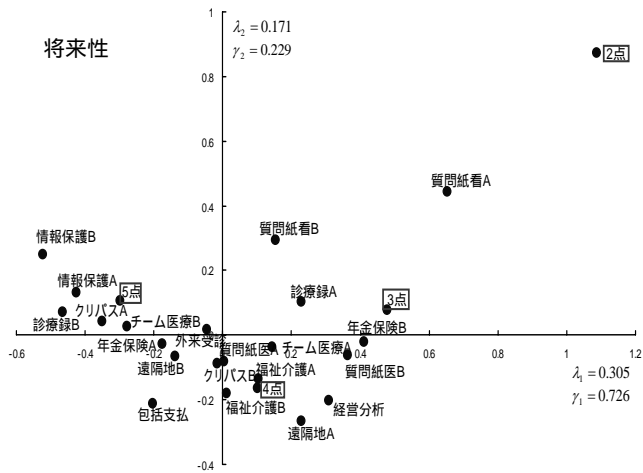
プレゼン



質問対応



将来性



データを解析する(10)

先ほどの2次元プロットで見たとき、どのグループが類
似しているのだろうか、そんな大きな枠組みを知りたい。
その類似具合を例えば偏差平方和で見てみよう。

そこで次ぎの図を見てほしい。

データを解析する(17)

この結果を積差率相関という尺度で示す。

各グループごとに示そう。

データを解析する(18)

		積差率相関	全変動
診療録 A	質問対応	0.7904	53.466
	将来性	0.8219	55.596
	プレゼン	0.8307	56.192
	勉強発表	0.8618	58.296
	内容適切	0.8801	59.529
	合計	4.1849	283.079
固有値		0.837	

Range=0.0897

		積差率相関	全変動
診療録 B	質問対応	0.6528	40.707
	将来性	0.7231	45.095
	内容適切	0.8129	50.693
	勉強発表	0.8591	53.578
	プレゼン	0.9091	56.694
	合計	3.957	246.767
固有値		0.791	

Range=0.2563

データを解析する(19)

		積差率相関	全変動
クリパス A	プレゼン	0.5563	32.961
	勉強発表	0.8428	49.935
	将来性	0.8464	50.146
	内容適切	0.8632	51.139
	質問対応	0.8913	52.806
	合計	4	236.987
	固有値		0.8

Range=0.335

		積差率相関	全変動
クリパス B	プレゼン	0.5554	34.6
	質問対応	0.794	49.465
	勉強発表	0.8215	51.1791
	将来性	0.8282	51.597
	内容適切	0.8476	52.805
	合計	3.8467	239.6461
	固有値		0.7693

Range=0.2922

データを解析する(20)

		積差率相関	全変動
遠隔地 A	質問対応	0.6104	39.234
	プレゼン	0.7301	46.925
	勉強発表	0.7944	51.061
	内容適切	0.9291	59.721
	将来性	0.9535	61.288
	合計	4.0175	258.229
固有値		0.8035	

Range=0.3431

		積差率相関	全変動
遠隔地 B	将来性	0.7606	51.054
	プレゼン	0.7811	52.434
	勉強発表	0.822	55.175
	質問対応	0.8511	57.133
	内容適切	0.9393	63.044
	合計	4.1541	278.84
固有値		0.8308	

Range=0.1787

データを解析する(21)

		積差率相関	全変動
情報保護 A	プレゼン	0.5147	32.415
	勉強発表	0.6417	40.408
	将来性	0.804	50.633
	質問対応	0.9506	59.862
	内容適切	0.9765	61.493
	合計	3.8875	244.811
固有値		0.7775	

Range=0.4618

		積差率相関	全変動
情報保護 B	プレゼン	0.7053	45.907
	内容適切	0.8148	53.036
	勉強発表	0.8197	53.356
	質問対応	0.8892	57.879
	将来性	0.891	57.996
	合計	4.12	268.174
固有値		0.824	

Range=0.1857

データを解析する(22)

		積差率相関	全変動
外来受診	内容適切	0.67	42.504
	プレゼン	0.7056	44.7626
	勉強発表	0.8528	54.099
	質問対応	0.8573	54.386
	将来性	0.8793	55.784
	合計	3.965	251.5356
固有値		0.793	

Range=0.2093

データを解析する(23)

		積差率相関	全変動
チム医療A	プレゼン	0.6679	42.498
	質問対応	0.7716	49.099
	将来性	0.7808	49.683
	内容適切	0.824	52.429
	勉強発表	0.8832	56.2
	合計	3.9275	249.909
	固有値	0.786	

Range=0.2153

		積差率相関	全変動
チム医療B	将来性	0.761	51.823
	プレゼン	0.7974	54.306
	内容適切	0.8446	57.52
	勉強発表	0.9147	62.289
	質問対応	0.9388	63.934
	合計	4.2565	289.872
	固有値	0.8513	

Range=0.1778

データを解析する(24)

		積差率相関	全変動
包括支払	勉強発表	0.693	44.975
	プレゼン	0.7619	49.446
	将来性	0.7711	50.042
	内容適切	0.8612	55.8892
	質問対応	0.9188	59.63
	合計	4.006	259.9822
	固有値	0.8012	

Range=0.2258

データを解析する(25)

		積差率相関	全変動
年金保険A	質問対応	0.7577	51.545
	内容適切	0.7909	53.805
	勉強発表	0.8649	58.84
	将来性	0.8751	59.531
	プレゼン	0.9109	61.967
	合計	4.1995	285.688
	固有値	0.8399	

Range=0.1532

		積差率相関	全変動
年金保険B	将来性	0.7304	46.843
	プレゼン	0.7513	48.184
	質問対応	0.8287	53.152
	勉強発表	0.8447	54.179
	内容適切	0.8739	56.049
	合計	4.029	258.407
	固有値	0.8058	

Range=0.1435

データを解析する(26)

		積差率相関	全変動
経営分析	勉強発表	0.7563	49.693
	将来性	0.7682	50.473
	質問対応	0.8284	54.432
	内容適切	0.8749	57.485
	プレゼン	0.8787	57.735
	合計	4.1065	269.818
	固有値	0.8213	

Range=0.1224

データを解析する(27)

		積差率相関	全変動
福祉介護A	内容適切	0.7158	48.311
	勉強発表	0.7651	51.637
	プレゼン	0.8741	58.999
	質問対応	0.8874	59.895
	将来性	0.9242	62.376
	合計	4.1666	281.218
	固有値	0.8333	

Range=0.2084

		積差率相関	全変動
福祉介護B	プレゼン	0.5923	38.212
	勉強発表	0.8311	53.619
	将来性	0.8624	55.64
	質問対応	0.878	56.648
	内容適切	0.8783	56.664
	合計	4.0421	260.783
	固有値	0.8084	

Range=0.2860

データを解析する(28)

		積差率相関	全変動
質問紙医A	プレゼン	0.4598	31.369
	質問対応	0.8909	60.775
	勉強発表	0.897	61.194
	将来性	0.9464	64.562
	内容適切	0.976	66.583
	合計	4.1701	284.483
	固有値	0.834	

Range=0.5162

		積差率相関	全変動
質問紙医B	勉強発表	0.6711	40.728
	将来性	0.7544	45.787
	内容適切	0.7609	46.18
	質問対応	0.7714	46.817
	プレゼン	0.8447	51.263
	合計	3.8025	230.775
	固有値	0.7605	

Range=0.1736

データを解析する(29)

	積差率相関	全変動
質問紙看A	プレゼン	47.835
	質問対応	51.43
	勉強発表	51.546
	将来性	54.514
	内容適切	61.948
	合計	267.273
固有値	0.8174	

Range=0.2158

	積差率相関	全変動
質問紙看B	プレゼン	39.999
	勉強発表	43.417
	質問対応	48.579
	内容適切	49.07
	将来性	52.433
	合計	233.498
固有値	0.763	

Range=0.2032

言いたいこと(1)

- 今回の発表会では皆さん比較的马ジメに評価していました。
- 今回の結果からグループの評価はできますが、個人は評価できません。
- 0点をつけた人が2名いましたが、1点と併合しても固有ベクトルの2次元図の位置関係は変わっていませんでした。

言いたいこと(2)

- 各グループと評価点の関係が概観できました。
- 皆さんは5つの質問項目で概ねできの良かったグループとその他のグループを選ぶ傾向があるようです。しかし、少し躊躇している様子も見えました。
- 評価が悪かった(良かった)グループでも、良かった(悪かった)点を積差率相関から見ることができました。

言いたいこと(3)

- 今回の皆さんの行動を数字で示すことができました。
- それには、皆さんが記述した質問紙という簡単なアンケートデータしか使っていません。
- 統計学という誰もが使える科学的な方法を用いたに過ぎません。

言いたいこと(4)

- 今後医学、医療が複雑になる一方で皆さんは、様々なデータという海を進まなければならないでしょう。
- おそらく、判別したり、主成分をとったり、似通っているものを分類したり、いろいろな作業を行うでしょう。
- そのとき、統計学を使ったアプローチをお願いしたいのです。それはデータを科学的に見ることへの足がかりとなるはずです。

結語

- 必要なデータの重要性を認識して、収集、その解析、それによって得られた結果の妥当性を確認し、考察するというプロセスを踏んでください。
- そのプロセスに統計学的な考えがあると、世界が違って見える(はず)と思います。

そして、少しいいからです、統計の基礎も勉強願います。