

質的データの数量化

昨年度の医療情報学実習期間中の変化を中心に

島根大学医学部医療情報学講座
安田 晃

今日の概要

- データって何だ.
- 質的って何だ.
- 質的データがあれば量的データもある.
- 類似していること.
- 類似していることを数字で示す.
- 類似していることをグラフで示す.
- グラフから何らかの隠れた構造が見えないか.
- これらのことから被験者集団の特性を定量的に表現したい.

データって何？

立証の材料として集められた、何らかの判断を導く情報を内包している事実。
日本国語大辞典, 小学館.

立証・計算の基礎となる、既知の、あるいは認容された事実・数値・資料・与件。
広辞苑第5版, 岩波書店.

一般に実験、観察、調査などによって得られた観察結果の資料をデータと呼ぶ。
統計用語辞典, 新曜社.

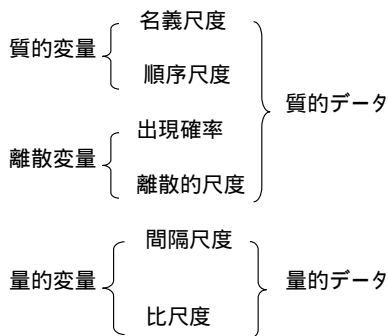
質的って何？

質に関するさま.

広辞苑第5版, 岩波書店.

質とは、生れつき、天性、内容、中味、価値、問い、ただすこと、飾りけがないこと。(quality)物がそれとして存在するもとであるもの。㊦対象を他の対象と区別する特色となっているもの。非感覚的な面も含む。「どのような」という問に対する事物のあり方。㊧論理学では、判断の肯定・否定をいう

データを分類する



質的変量 その1

名義尺度・・・測定対象をいくつかのカテゴリに分類するだけの尺度.

例 あなたは昼食に何を食べる予定ですか？

- 1 おかず2品の定食(450円)
- 2 チキン南蛮弁当(490円)
- 3 割子そば(うずらの卵入り)(280円)
- 4 ラーメン(190円)
- 5 その他(できるなら500円まで)

この場合、選択肢の順序は問題ではない

質的変数 その2

順序尺度・・・測定対象を順序性のあるカテゴリに分類する尺度.

例 今日,あなたが昼食に使う予算は?

- 1 ~ 250円
- 2 251円 ~ 450円
- 3 451円 ~ 500円
- 4 501円を超えるがせいぜい750円まで

この場合,順序が問題になる

量的変数 その1

間隔尺度・・・測定対象の量の大きさを測定値間の数値の差のみで示す尺度.

例 タングステン(W)の沸点5700 は窒素(N₂)のそれ-198 より5898 高い.20点のAさんは70点のBさんより50点悪い.

この場合,差だけが問題となる.決して $5700/|-198| = 28.79$ 倍高いとは言わない.

量的変数 その2

比尺度・・・測定値の原点が一義的に決まっている尺度.

例 出雲市の人口88591人は松江市の人口148987人と比べ,60396人少ない.あるいは,米子市のそれと比べ0.6286倍である.
地球の脱出速度11.2km/sは120km/hで走った伯備線やくも号より336.02倍速い(当然やくも号では宇宙旅行は出来ない).

この場合,差だけではなく,測定値の倍数関係も議論できる.

今回用いるデータは・・・

2003年度後期医療情報学実習で毎週にわたって記述してもらった医療情報学実習評価シート

医療情報学実習評価シート()

関連事項の想起と学習項目の抽出

- | |
|--------------------------------------|
| 1) 多目的な発送や統合的な連想ができたか(発送連想) |
| 2) 課題に含まれる重要なテーマに気づいたか(重要なテーマ) |
| 3) 既に学んだ知識の整理ができたか(知識の整理) |
| 4) 課題からさまざまな疑問点や学習項目を抽出できたか(学習項目の抽出) |

学習対象の取捨選択

- | |
|---|
| 5) 抽出した学習項目を重要度にしたがって順位づけできたか(学習項目順位付け) |
| 6) グループ全員に共通な学習項目を設定できたか(共通な学習項目) |
| 7) 自分独自の学習項目を設定できたか(独自の学習項目) |
| 8) 基本的な事項を学ぼうとしたか(基本的事項) |
| 9) 発展的・応用的な事項を学ぼうとしたか(発展応用) |

医療情報学実習評価シート()

学習計画の立案と実行

- | |
|------------------------------------|
| 10) 学習目標ごとに自らの到達目標を設定できたか(発送連想) |
| 11) 学習計画の時間配分は適切であったか(学習時間の時間配分) |
| 12) 問題を解決するために具体的な方法を見い出せたか(具体的方法) |
| 13) 自己学習に十分な時間と努力を注いだか(自己学習) |
| 14) 自己が設定した到達目標を達成できたか(到達目標) |

グループ学習上の態度

- | |
|---------------------------------------|
| 15) 自分の考えを簡明かつ論理的に説明できたか(論理的説明) |
| 16) 他者の考えを理解しようと努めたか(他者の理解) |
| 17) 自分の考えと異なる意見に対しても柔軟な態度がとれたか(柔軟な態度) |
| 18) 討論や発表時間に留意したか(討論発表の時間配分) |
| 19) グループの一員として問題解決への建設的な貢献をしたか(建設的貢献) |

(今回使う)類似していることとは,
たとえば...

あなたは次の食べ物が好きですか？

1. そば
2. うどん
3. 鯖の味噌煮
4. おでん
5. ビスケット
6. ポテトサラダ
7. 野菜いため

このようなデータが得られたとする

被験者 No.	そば	うどん	さばの味噌煮	おでん	ビスケット	ポテトサラダ	野菜炒め
1							
2							
3							
4							
5							
6							
7							
8							
9							
10							

この表から

そばとうどん両方選好した人は5名,
そばとさばの味噌煮両方選好した人は5名,

⋮

おでんとビスケット両方選好した人は2名,

⋮

ポテトサラダと野菜炒め両方選好した人は1名.

このようにも書けるだろう

	そば	うどん	さばの味噌煮	おでん	ビスケット	ポテトサラダ	野菜炒め
そば	-						
うどん	5	-					
さばの味噌煮	5	7	-				
おでん	5	7	6	-			
ビスケット	2	2	2	2	-		
ポテトサラダ	3	4	3	3	1	-	
野菜炒め	3	6	5	5	2	1	-

あるいは、非選好という指標から

	そば	うどん	さばの味噌煮	おでん	ビスケット	ポテトサラダ	野菜炒め
そば	-						
うどん	5	-					
さばの味噌煮	5	3	-				
おでん	5	3	4	-			
ビスケット	8	8	8	8	-		
ポテトサラダ	7	6	7	7	9	-	
野菜炒め	7	4	5	5	8	9	-

7種類の食品を(非)選好度という(非)類似性で見た場合、
このような表が書けた

そこで、似ているものは小さい数字
で、似ていないものは大きな数字と
なるように関数を作ってみよう。

一般化して、

今、個体 i の p 次元ベクトルを

$$\mathbf{x}_i = (x_{i1}, \dots, x_{ip})^T \quad (i = 1, \dots, n)$$

と表わす。そのとき、個体 i と個体 j のユークリッド距離は

$$\begin{aligned} d_{ij}^2 &= \sum_{k=1}^p (x_{ik} - x_{jk})^2 = \|\mathbf{x}_i - \mathbf{x}_j\|^2 \\ &= \mathbf{x}_i^T \mathbf{x}_i - 2\mathbf{x}_i^T \mathbf{x}_j + \mathbf{x}_j^T \mathbf{x}_j \end{aligned}$$

と書ける。

ここで、この n 行 p 列のユークリッド距離で示した
行列は、各要素から平均値を引いた平均偏差得点
行列 X においても変わらない。よって (i, j) 成分を
要素とする距離行列を、

$$D^2 = \begin{bmatrix} 0 & d_{12}^2 & \cdots & d_{1n}^2 \\ d_{21}^2 & 0 & \cdots & d_{2n}^2 \\ \vdots & \vdots & \ddots & \vdots \\ d_{n1}^2 & d_{n1}^2 & \cdots & 0 \end{bmatrix}$$

とおくと、

$$D^2 = \mathbf{I}_n \mathbf{I}_n^T \text{diag}(XX^T) - 2XX^T + \text{diag}(XX^T) \mathbf{I}_n \mathbf{I}_n^T$$

と書ける。ここで、

$$P_M^\perp \mathbf{I}_n = \left(\mathbf{I}_n - \frac{1}{n} \mathbf{I}_n \mathbf{I}_n^T \right) \mathbf{I}_n = 0$$

を用いて、新たな行列は

$$H = -\frac{1}{2} P_M^\perp D P_M^\perp = XX^T$$

となる。この H は非負定値行列である

行列 $H = (h_{ij})$ の h_{ij} 成分は、

$$h_{ij} = (-1/2)(d_{ij}^2 - \bar{d}_{i\cdot}^2 - \bar{d}_{\cdot j}^2 + d_{\cdot\cdot}^2)$$

ここで、 $\bar{d}_{i\cdot}^2 = (1/n) \sum_j d_{ij}^2$, $\bar{d}_{\cdot j}^2 = (1/n) \sum_i d_{ij}^2$

$$d_{\cdot\cdot}^2 = (1/n) \sum_i \sum_j d_{ij}^2$$

となる。これらの変換を、ヤング・ハウスホルダ変換

といって、(非)類似性を定量的に調べる場合重要

となる。

これまでの準備から

医療情報学評価シートを考える。

用いたデータは,

毎実習終了時に記入した

妥当であった否かを答えた.

解析の目的は,

どの期間,どの質問項目を選好したか.

どの程度類似しているか

どの程度非類似しているか

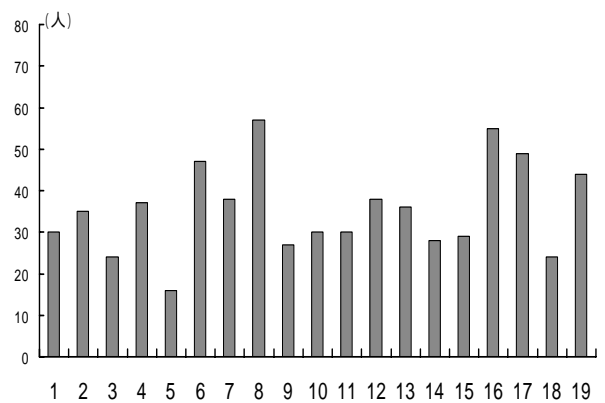
解析の方法は,

各週記載した評価シートの19項目の非類似性を調べる.

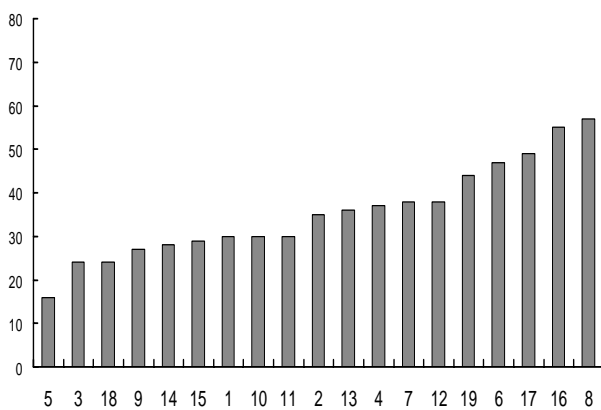
19×19の要素をもつ行列をつくる.ここで対角要素は0とおく.

10月10日から12月5日までの7週それぞれを交換不可能な順序ある個々のデータとして解析する.

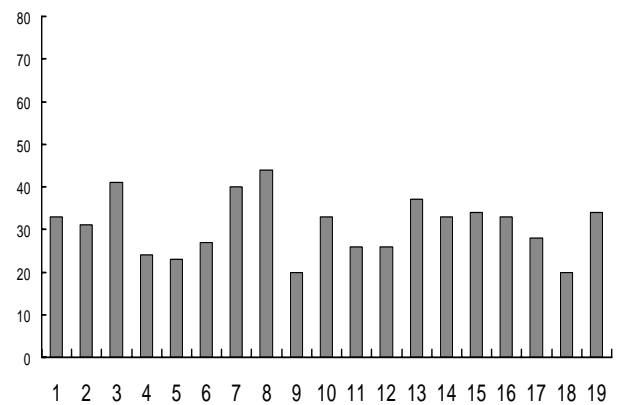
10月10日のデータ



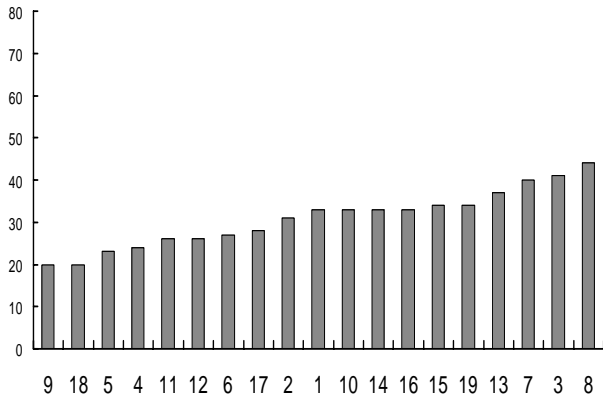
10月10日のデータ



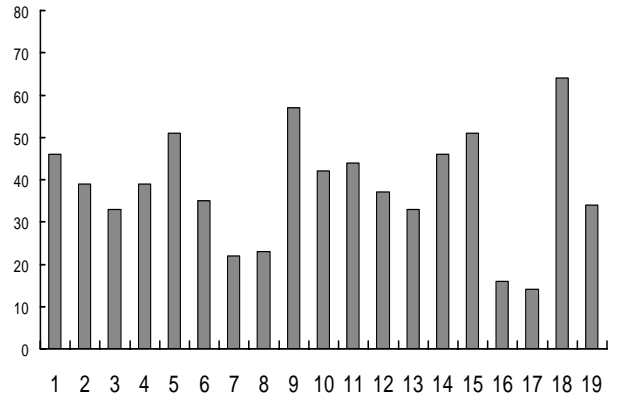
10月24日のデータ



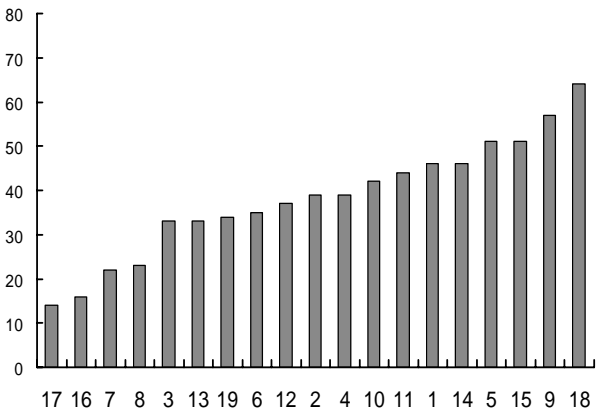
10月24日のデータ



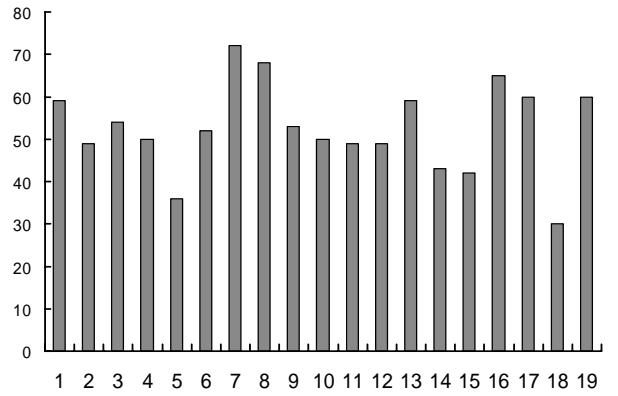
10月31日のデータ



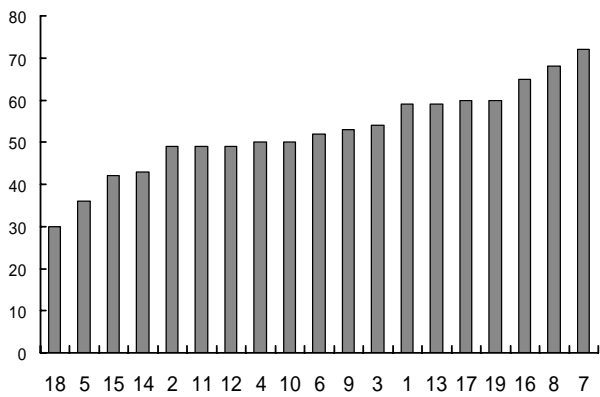
10月31日のデータ



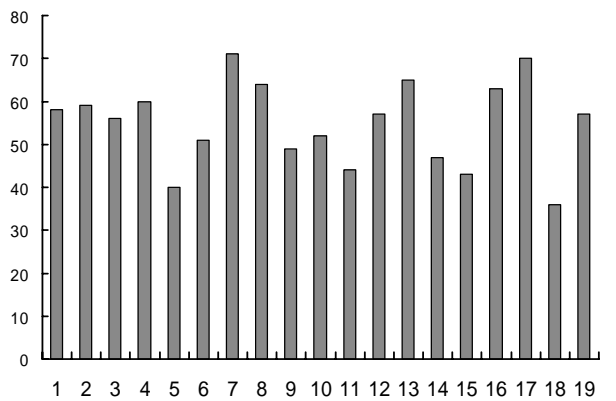
11月7日のデータ



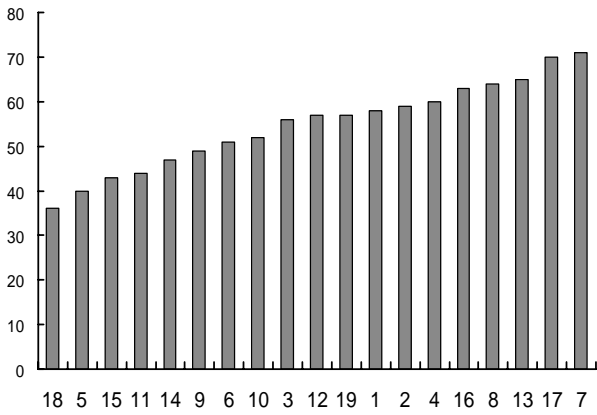
11月7日のデータ



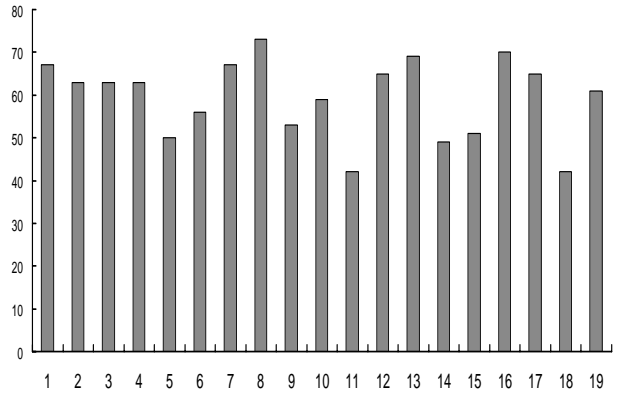
11月14日のデータ



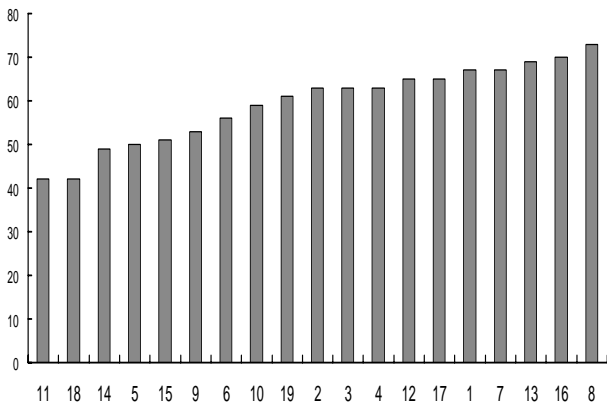
11月14日のデータ



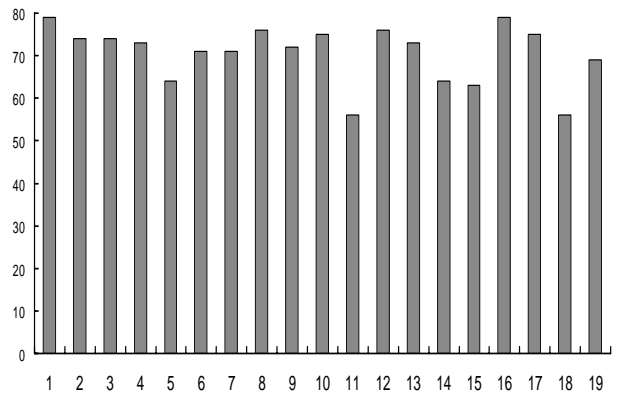
11月28日のデータ



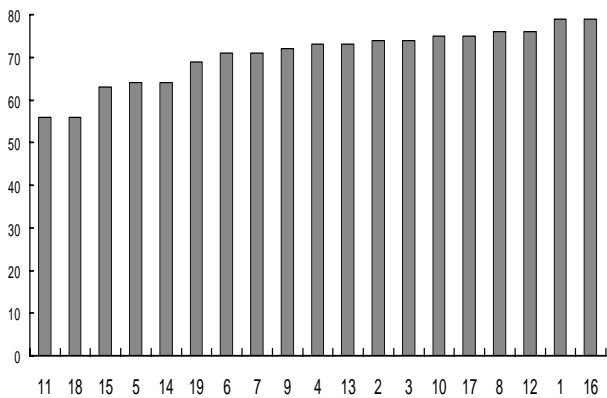
11月28日のデータ



12月5日のデータ



12月5日のデータ



10月10日の類似度行列(下三角行列)

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	
1) 発想連想	0																			
2) 重要なテーマ	20	0																		
3) 知識の整理	12	14	0																	
4) 学習項目の抽出	20	22	13	0																
5) 学習項目間の付加	13	11	9	12	0															
6) 共通な学習項目	22	25	17	24	12	0														
7) 独自の学習項目	22	23	14	22	11	26	0													
8) 基本的事項	25	26	22	32	14	34	30	0												
9) 発展内容	19	18	11	18	13	20	17	25	0											
10) 学習目標	17	20	14	18	12	25	18	23	17	0										
11) 学習時間の時間配分	17	20	17	17	11	23	20	26	16	19	0									
12) 具体的方法	20	22	15	20	10	27	24	32	18	16	19	0								
13) 自己学習	21	22	13	22	13	25	24	29	21	17	21	22	0							
14) 到達目標	18	19	13	23	11	20	20	24	15	16	16	14	19	0						
15) 論理的思考	20	20	13	20	13	22	19	22	16	16	15	18	18	16	0					
16) 他者の理解	24	31	19	30	15	37	28	44	24	25	27	30	28	23	24	0				
17) 柔軟な態度	25	26	18	29	15	36	28	38	24	25	24	25	26	25	24	42	0			
18) 討論発表の時間配分	15	17	11	16	10	19	17	20	15	14	18	17	19	14	17	21	20	0		
19) 建設的意見	23	24	19	25	15	32	24	36	21	20	22	24	24	22	23	36	36	18	0	

10月24日の類似度行列

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
1) 発想連想	0																		
2) 重要なテーマ	16	0																	
3) 知識の整理	22	18	0																
4) 学習項目の抽出	16	17	17	0															
5) 学習項目順位付け	21	14	18	14	0														
6) 共通な学習項目	19	15	18	17	13	0													
7) 独自の学習項目	26	23	26	17	19	18	0												
8) 基本的事項	22	21	28	21	17	17	24	0											
9) 発展応用	14	14	17	10	12	12	15	14	0										
10) 学習目標	21	18	22	15	18	15	24	20	13	0									
11) 学習時間の時間配分	19	15	17	15	13	15	19	18	11	16	0								
12) 具体的方法	17	16	18	12	15	16	20	18	12	15	12	0							
13) 自己学習	21	19	23	19	16	19	24	26	15	19	17	15	0						
14) 到達目標	23	17	24	15	18	16	21	23	13	20	15	15	17	0					
15) 論理的説明	25	21	24	18	19	20	27	24	14	21	18	20	24	20	0				
16) 他者の理解	21	18	20	17	14	16	24	22	15	17	17	13	20	20	19	0			
17) 柔軟な態度	19	16	20	18	16	15	19	22	12	16	16	13	18	16	18	20	0		
18) 討議発表の時間配分	16	12	18	11	12	15	16	15	15	13	13	11	17	13	17	15	13	0	
19) 建設的意見	24	19	20	18	17	19	22	24	13	21	17	17	25	20	23	22	16	13	0

10月31日の類似度行列

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
1) 発想連想	0																		
2) 重要なテーマ	30	0																	
3) 知識の整理	26	20	0																
4) 学習項目の抽出	27	26	23	0															
5) 学習項目順位付け	36	30	25	28	0														
6) 共通な学習項目	23	20	17	18	28	0													
7) 独自の学習項目	17	13	12	13	17	11	0												
8) 基本的事項	18	15	14	16	16	8	9	0											
9) 発展応用	36	27	23	31	40	27	19	17	0										
10) 学習目標	26	24	23	26	30	24	20	14	33	0									
11) 学習時間の時間配分	30	26	21	26	34	22	14	12	35	27	0								
12) 具体的方法	23	27	15	23	28	21	14	15	28	24	22	0							
13) 自己学習	23	17	20	20	22	17	13	11	28	23	23	19	0						
14) 到達目標	31	26	22	27	36	27	13	15	36	28	29	27	22	0					
15) 論理的説明	36	29	24	28	38	28	16	18	41	31	31	33	26	36	0				
16) 他者の理解	13	13	8	9	12	13	5	8	12	10	10	13	8	11	13	0			
17) 柔軟な態度	12	10	6	6	12	11	4	5	12	8	9	11	5	11	14	10	0		
18) 討議発表の時間配分	41	36	30	34	45	33	17	20	49	37	42	33	32	42	47	16	14	0	
19) 建設的意見	29	22	19	20	28	20	14	13	28	21	25	23	21	25	30	14	12	31	0

11月7日の類似度行列

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
1) 発想連想	0																		
2) 重要なテーマ	38	0																	
3) 知識の整理	40	33	0																
4) 学習項目の抽出	41	32	38	0															
5) 学習項目順位付け	27	25	26	23	0														
6) 共通な学習項目	42	35	37	34	21	0													
7) 独自の学習項目	52	46	50	46	33	47	0												
8) 基本的事項	51	43	47	41	30	43	60	0											
9) 発展応用	42	36	38	35	26	38	45	44	0										
10) 学習目標	41	34	39	38	22	39	47	43	34	0									
11) 学習時間の時間配分	36	32	34	31	25	35	44	43	33	32	0								
12) 具体的方法	35	35	33	32	23	34	46	42	31	34	34	0							
13) 自己学習	45	39	41	41	29	40	53	51	40	39	40	38	0						
14) 到達目標	35	29	29	29	22	36	40	37	30	32	32	27	31	0					
15) 論理的説明	36	30	28	30	22	32	37	32	32	31	26	25	32	28	0				
16) 他者の理解	47	41	47	44	27	45	59	57	45	43	41	42	51	35	34	0			
17) 柔軟な態度	47	38	41	42	26	43	55	51	39	42	40	39	47	33	33	52	0		
18) 討議発表の時間配分	23	24	21	22	20	21	30	25	21	22	25	25	27	22	19	26	23	0	
19) 建設的意見	47	42	43	40	28	45	54	51	43	43	37	40	44	37	36	52	49	26	0

11月14日の類似度行列

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
1) 発想連想	0																		
2) 重要なテーマ	50	0																	
3) 知識の整理	46	45	0																
4) 学習項目の抽出	49	51	44	0															
5) 学習項目順位付け	33	34	32	33	0														
6) 共通な学習項目	41	42	42	42	30	0													
7) 独自の学習項目	55	57	54	56	39	49	0												
8) 基本的事項	50	54	48	54	37	46	60	0											
9) 発展応用	43	42	41	41	29	39	46	41	0										
10) 学習目標	43	45	39	43	32	38	50	47	37	0									
11) 学習時間の時間配分	37	36	39	38	28	35	42	38	35	32	0								
12) 具体的方法	45	47	45	47	32	44	56	53	39	44	35	0							
13) 自己学習	49	54	48	52	38	46	61	56	45	47	40	50	0						
14) 到達目標	40	39	40	39	28	34	46	41	36	36	35	38	44	0					
15) 論理的説明	36	37	35	37	29	36	43	38	30	37	29	36	41	29	0				
16) 他者の理解	49	51	47	51	38	46	60	58	40	46	38	52	55	40	39	0			
17) 柔軟な態度	54	56	54	55	39	50	69	60	46	50	42	56	60	46	43	60	0		
18) 討議発表の時間配分	31	32	29	31	24	27	33	32	31	30	29	29	33	28	23	31	34	0	
19) 建設的意見	48	47	44	49	33	40	56	51	41	44	35	47	51	39	37	50	55	31	0

11月28日の類似度行列

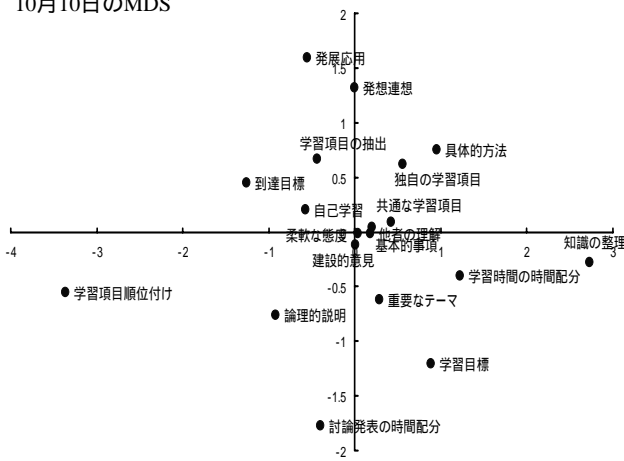
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
1) 発想連想	0																		
2) 重要なテーマ	55	0																	
3) 知識の整理	56	50	0																
4) 学習項目の抽出	54	53	53	0															
5) 学習項目順位付け	46	42	41	41	0														
6) 共通な学習項目	51	45	48	46	38	0													
7) 独自の学習項目	56	55	54	57	42	47	0												
8) 基本的事項	64	59	60	61	49	53	62	0											
9) 発展応用	47	46	44	43	40	42	45	51	0										
10) 学習目標	53	49	50	49	46	42	48	56	42	0									
11) 学習時間の時間配分	37	33	38	38	30	35	38	39	27	36	0								
12) 具体的方法	56	53	55	55	43	49	58	61	45	47	36	0							
13) 自己学習	58	58	56	58	46	50	60	64	50	54	39	56	0						
14) 到達目標	43	41	39	40	33	38	42	45	34	37	31	44	43	0					
15) 論理的説明	46	43	43	44	40	41	42	50	37	45	33	44	46	33	0				
16) 他者の理解	59	57	57	60	45	49	63	67	48	54	39	60	64	43	46	0			

今までの類似度行列の各要素を回答者数で引き算し、非類似度を計算する。つまりこの数字が大きいほど非類似している。

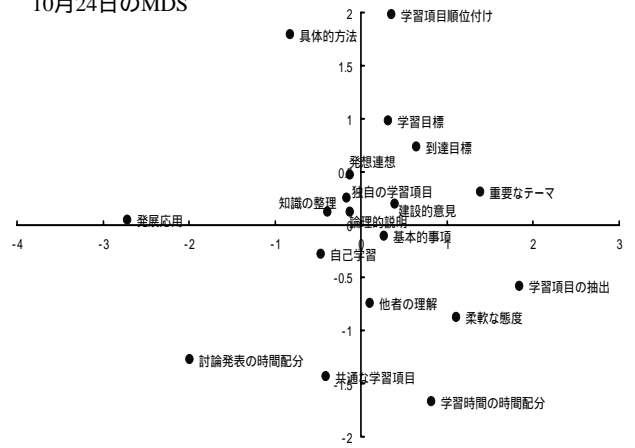
そして、実習した日の差異で非類似性を見てみたい。つまり7日間の実習で得られた7つの非類似度行列を、実習日を基準に固有ベクトルを2次元プロットしてみよう。

このように非類似度に着目して、非類似度が与えられたときに、その大小関係を個体間の距離が反映するように、各個体間を空間内に布置する解析方法を多次元尺度構成法 (Multidimensional scaling, MDS) という。

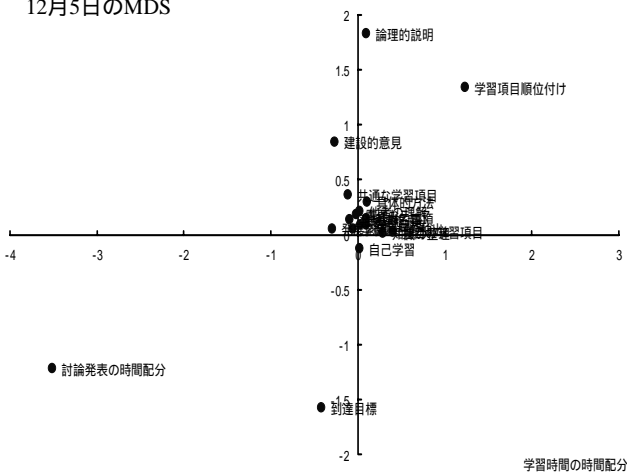
10月10日のMDS



10月24日のMDS



12月5日のMDS



今までのグラフから、

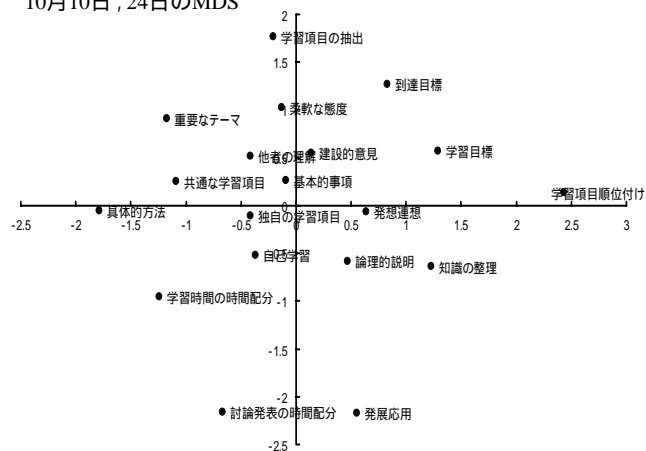
12月5日は非類似度(類似度)の小さな(大きな)項目が集まっている。

一方、実習の初期では19項目がかなりばらつきている。互いの類似性が最終週より見られない。

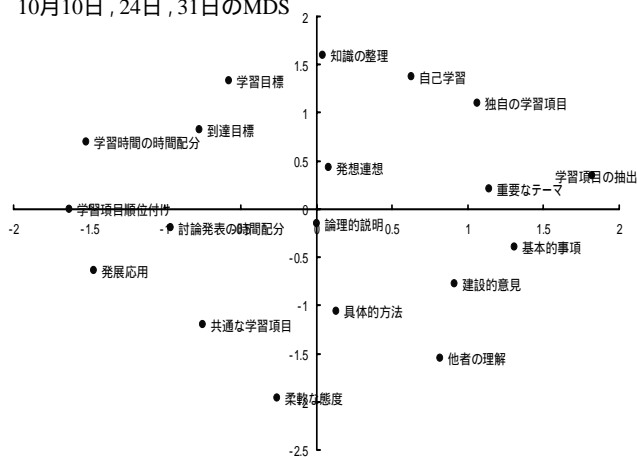
そこで、7週のそれぞれの週を集め、その集団でMDSを行ってみる。

例えば1週から2週まで、3週から7週までと。ここで、例えば1, 3, 4週とその他とか、1, 6週と2, 3, 4, 5, 7週という順序を無視した選択はできない。

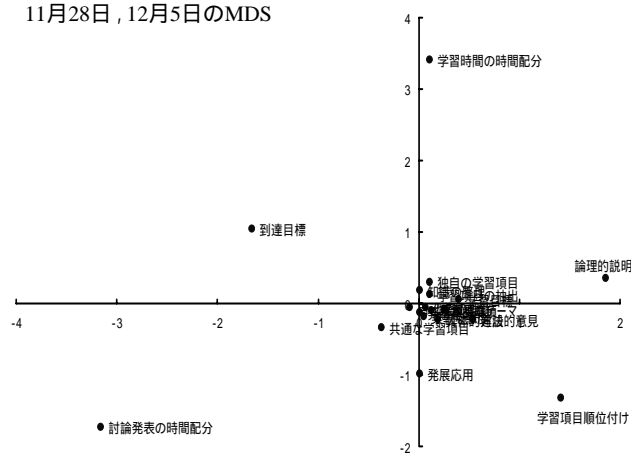
10月10日, 24日のMDS



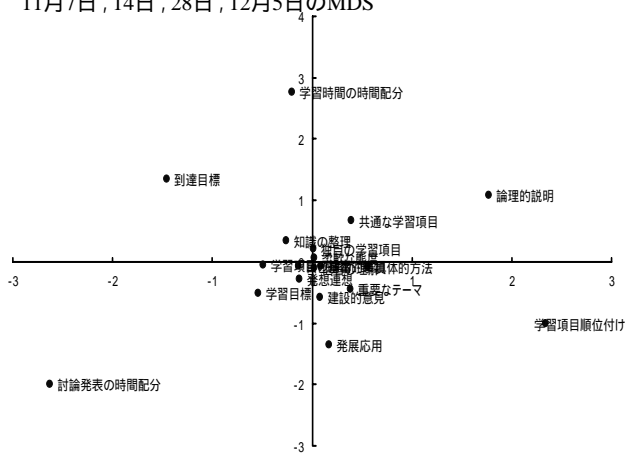
10月10日, 24日, 31日のMDS



11月28日, 12月5日のMDS



11月7日, 14日, 28日, 12月5日のMDS



以上より,

1. 実習前半と後半では意識が異なっているように見える。
2. 後半では非類似度が著しく小さく、5つの項目を除いて似たような意識で取り組んでいたと思われる。
3. 前半は意識が散逸しているようにも思われる。

前半後半で意識の違いが明らかになっているが、なぜこのような結果になったのだろうか。

…前後半の間に何らかの意識を変えるものが存在したであろう。それはおそらく…、と思われる。

まとめ

1. 医療情報学実習評価シートの各項目の非類似性に注目した。
2. 非類似度を関数とした非負定値行列をつくった。
3. その行列を多次元尺度構成法で解析した。
4. それによって、皆さんの医療情報学実習に対する取り組みの一部が定量的にわかった。