Концепция скрытых (латентных) переменных в химическом анализе. Часть 1. Качественный анализ

Померанцев Алексей Леонидович



Институт химической физики РАН им Семенова



Российское хемометрическое общество

Вопрос 1. Кто я?

Еще студент

Уже не студент

Вопрос 2. Моя направленность

Экспериментатор

Теоретик

Я ошибся комнатой

Вопрос 3. Специальность

Химик Физик Биолог Математик

Вопрос 4. Область

Аналитическая химия

Органическая химия

Физическая химия

Не занимаюсь ничем подобным

Вопрос 5. Специализация

Электрохимия

Хроматография

Спектроскопия

Масспектрометрия

Хемометрика

Вопрос 6. Знаю (читал) ...

1 журнал по аналитической химии

2 журнала по АХ

4 журнала по АХ

больше 6 журналов по АХ

Вопрос 7. Хемометрика это ...

статистика в аналитической химии

метрология применительно к АХ

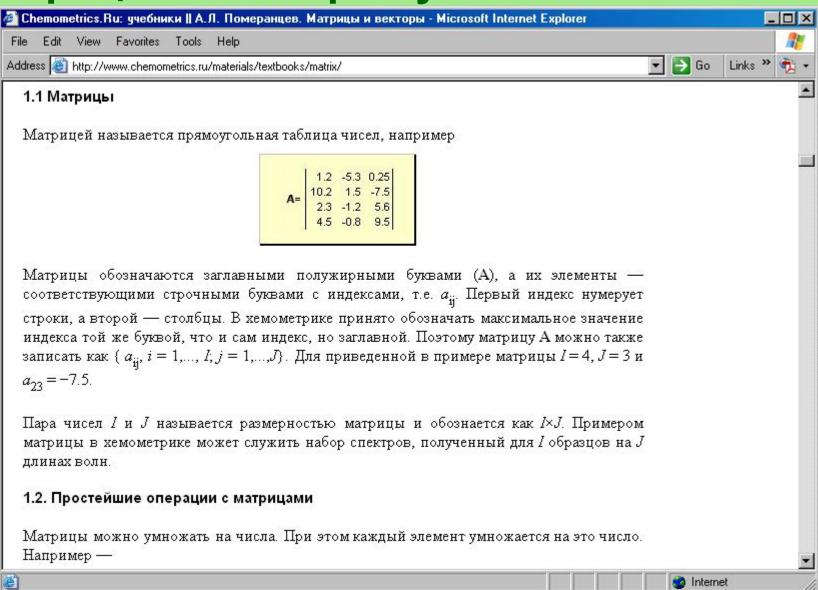
то, что делают хемометрики

понятия не имею ?!

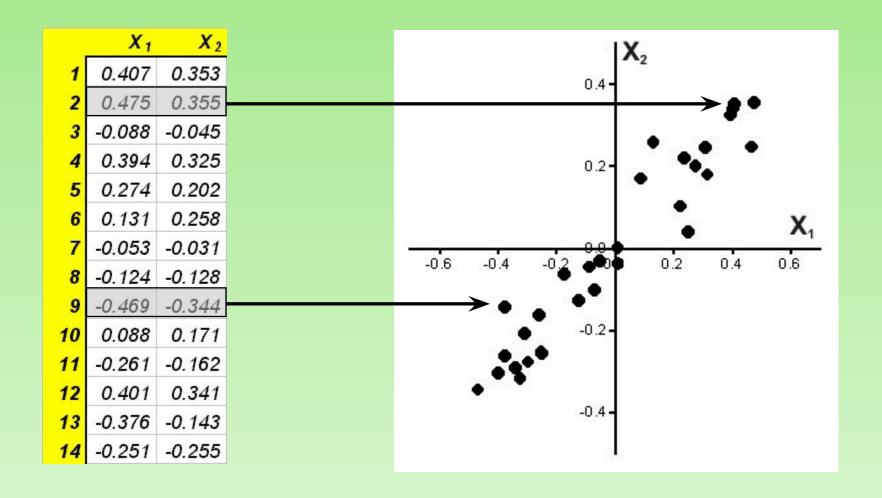
Данные о нас

Персона	Студент	Экс/Теор	Наука	Область	Специал	Журнал	Хемом
Иванов И.	1	1	1	1	1	1	1
Петров П.	1	1	1	1	1	1	1
Сидоров С.	1	1	обра	зец/об	ъект	2	2
	0	0	⊐	1	3	1	1
	1	1	ep	1	1	1	1
	0	1	e	1	4	1	1
	0	1	Т переменна	1	5	1	3
	0	1	Ĭκα	1	1	1	1
	1	0	ivi	атриі	4a ₁	2	4
	0	1	7R	аннь	1	1	1
Зонтов Ю.	1	0	B A	аппр	6	0	1
	0	1	Q	1	2	1	1
	0	0	СВОЙСТВО	1	3	1	2
	1	1	ГB	1	1	1	2
Померанцев А.	0	0	9	1	6	12	3
	1	1	1	2	2	1	1
	1	1	1	1	1	2	3
	1	1	2	1	2	1	2

Матрицы и векторы: учебник



Графическое представление 2D-данных



Графическое представление 3D-данных

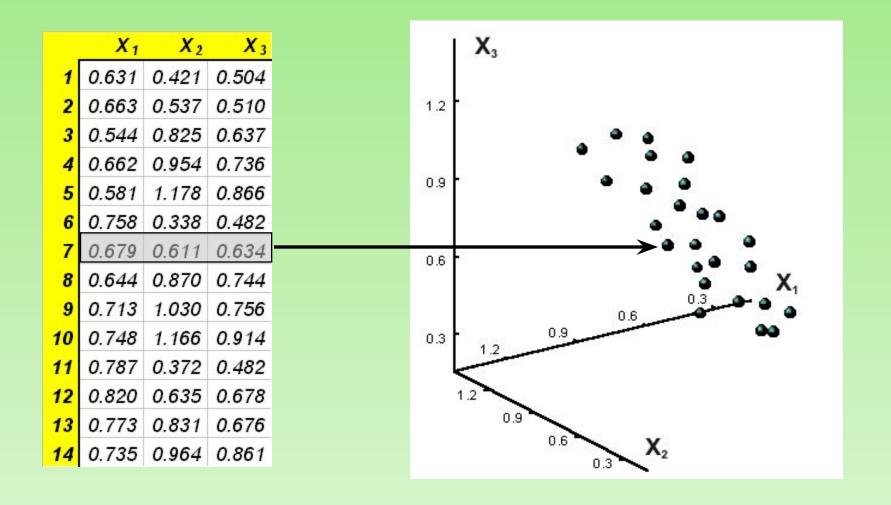


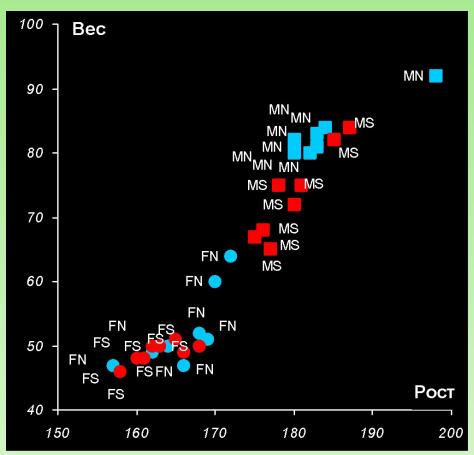


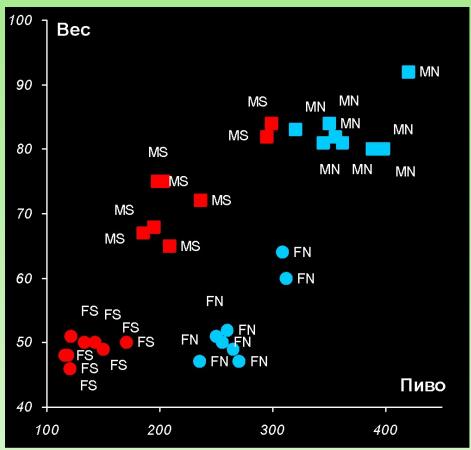
Таблица данных

Ī	Терсона	Пол	Регион	Рост	Bec	Волосы	Обувь	Возраст	Доход	Пиво	Вино	Сила	IQ
	Nº	M/F	S/N	СМ	КГ	S/L	размер	лет	К€/год	л/год	л/год	баллы	баллы
	1	M	N	198	92	S	48	48	45	420	115	98	100
	2	M	N	184	84	S	44	33	33	350	102	92	130
	3	M	N	183	83	S	44	37	34	320	98	91	127
	4	M	N	182	80	S	42	35	30	398	65	85	140
	5	M	N	180	80	S	43	36	30	388	63	84	129
	6	M	N	183	81	S	42	37	35	345	45	90	105
	7	M	N	180	82	S	44	43	37	355	82	88	109
	8	M	N	180	81	S	44	46	42	362	90	86	113
	9	M	S	185	82	S	45	26	16	295	180	92	109
	10	M	S	187	84	S	46	27	17	299	178	95	119
	11	M	S	177	65	S	41	26	18	209	160	86	120
	12	M	S	180	72	S	43	33	19	236	175	85	115
	13	M	S	181	75	S	43	42	31	198	161	83	105
	14	M	S	176	68	S	42	50	36	195	177	82	96
	15	M	S	175	67	L	42	55	38	185	187	80	105
	16	M	S	178	75	S	42	30	24	203	208	81	118
	17	F	N	166	47	S	36	32	28	270	78	75	112
	18	F	N	170	60	L	38	23	20	312	99	81	110
	19	F	N	172	64	L	39	24	22	308	91	82	102
	20	F	N	169	51	L	36	24	23	250	89	78	98
	21	F	N	168	52	L	37	27	24	260	86	78	100
	22	F	N	157	47	L	36	32	32	235	92	70	127
	23	F	N	164	50	L	38	41	34	255	134	76	101
	24	F	N	162	49	L	37	40	34	265	124	75	108
	25	F	S	168	50	L	37	49	34	170	162	76	135
	26	F	S	166	49	L	36	21	14	150	245	75	123
	27	F	S	158	46	L	34	30	18	120	120	70	119
	28	F	S	163	50	L	36	18	11	143	136	75	102
	29	F	S	162	50	L	36	20	12	133	146	74	132
	30	F	S	165	51	L	36	36	26	121	129	76	126
	31	F	S	161	48	L	35	41	32	116	196	75	120
٠,	32	F	S	160	48	L	35	40	31	118	198	74	129

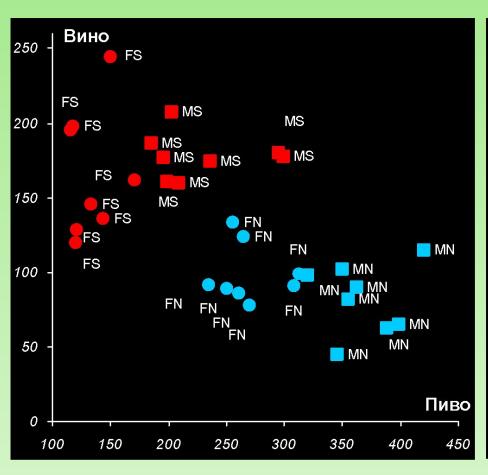
13.03.07

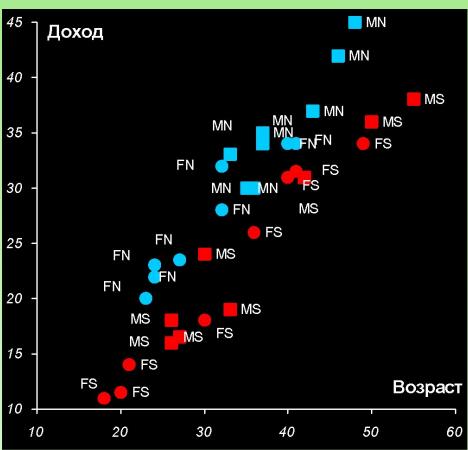
Корреляции 1



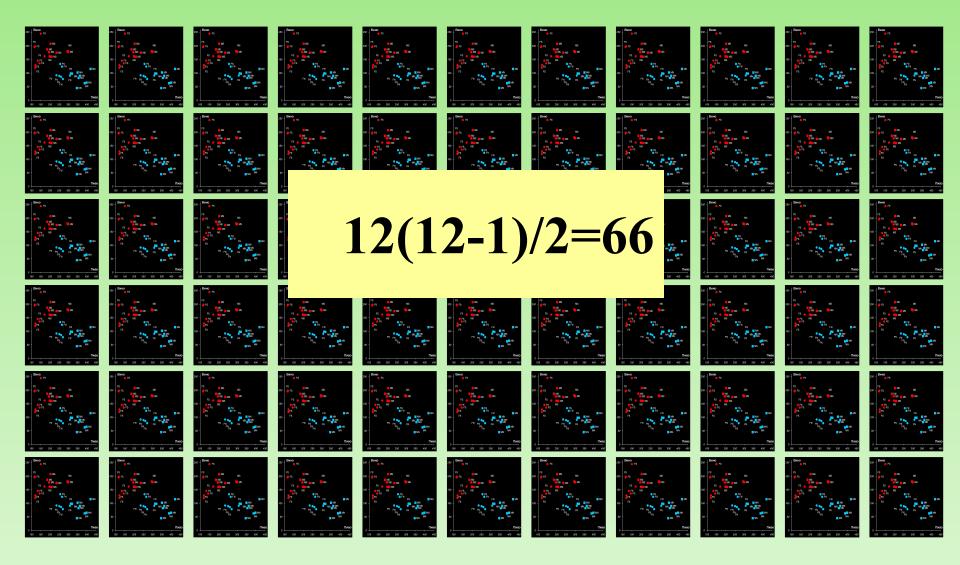


Корреляции 2





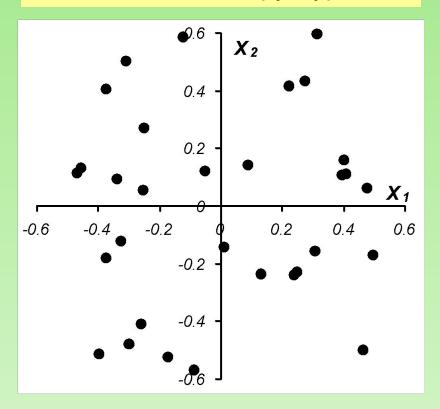
Все возможные корреляции



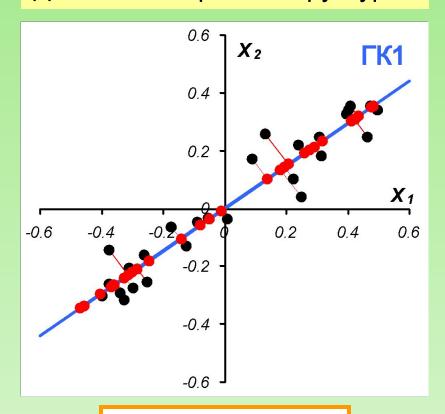
13.03.07

Главные направления и проекции

Данные без структуры

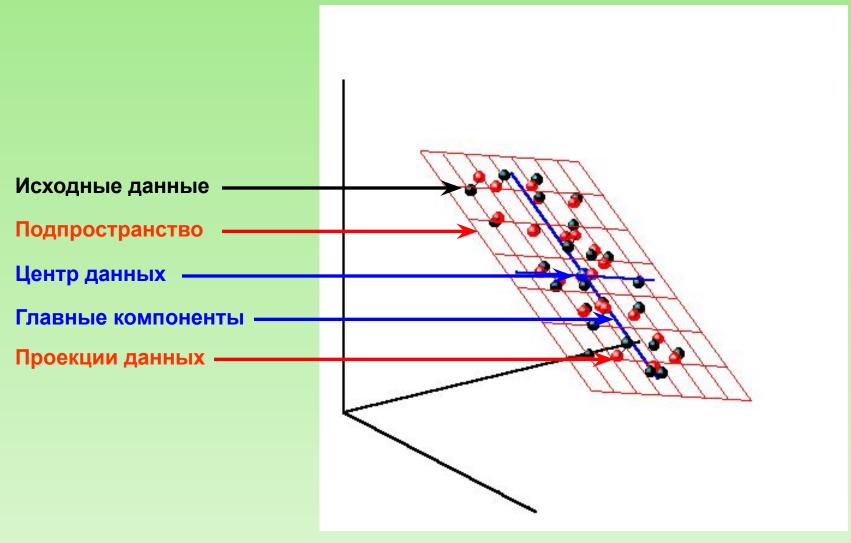


Данные со скрытой структурой



$$\mathbf{X}_2 = \mathbf{a}\mathbf{X}_1 + \mathbf{E}$$

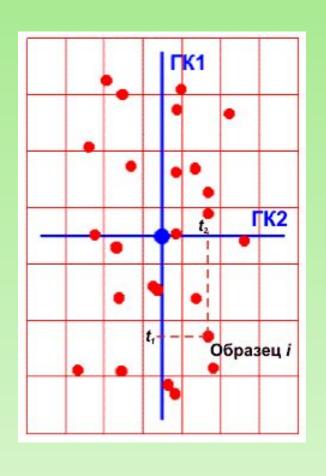
Проекция на подпространство

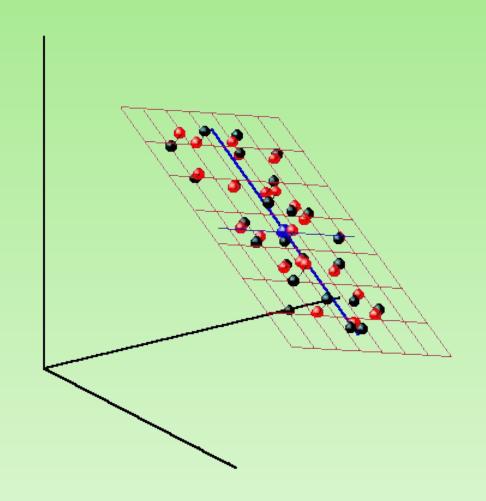


13.03.07

Лекция в МГУ

Представление данных в подпространстве

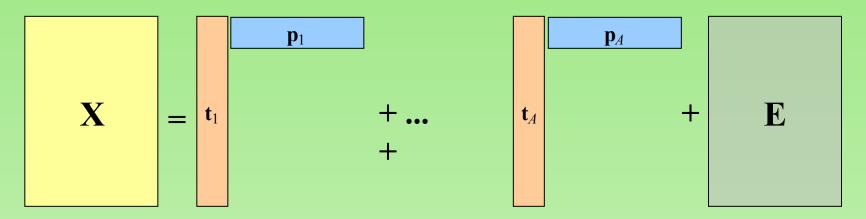




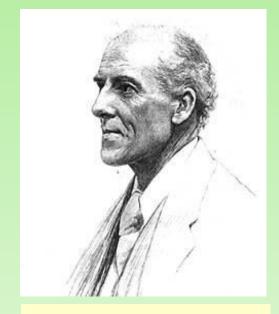
13.03.07

Лекция в МГУ

Метод главных компонент: МГК (РСА)

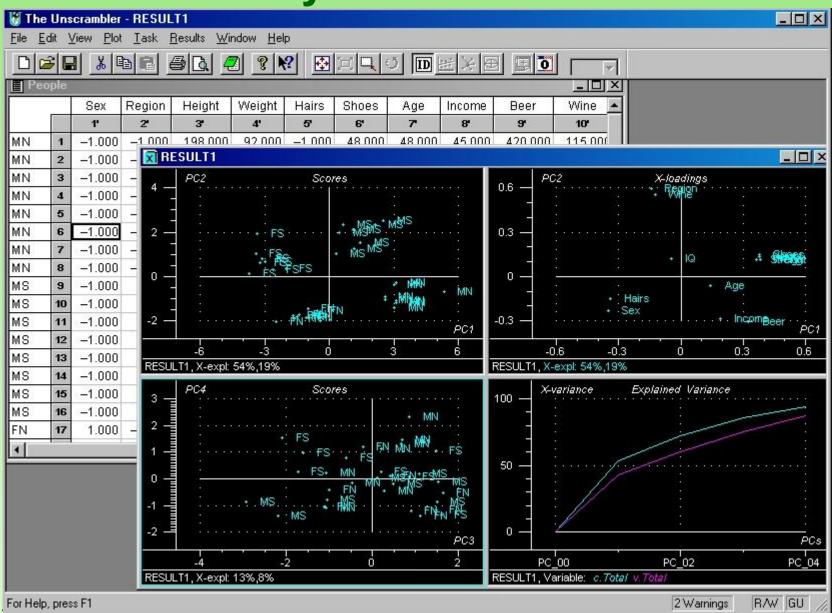


- Х матрица данных,
- **E** матрица погрешностей, обе $(I \times J)$
- **T** матрица *счетов*: $(I \times A)$,
- **Р** матрица *нагрузок*: $(A \times J)$
- A число главных компонент (A << J)

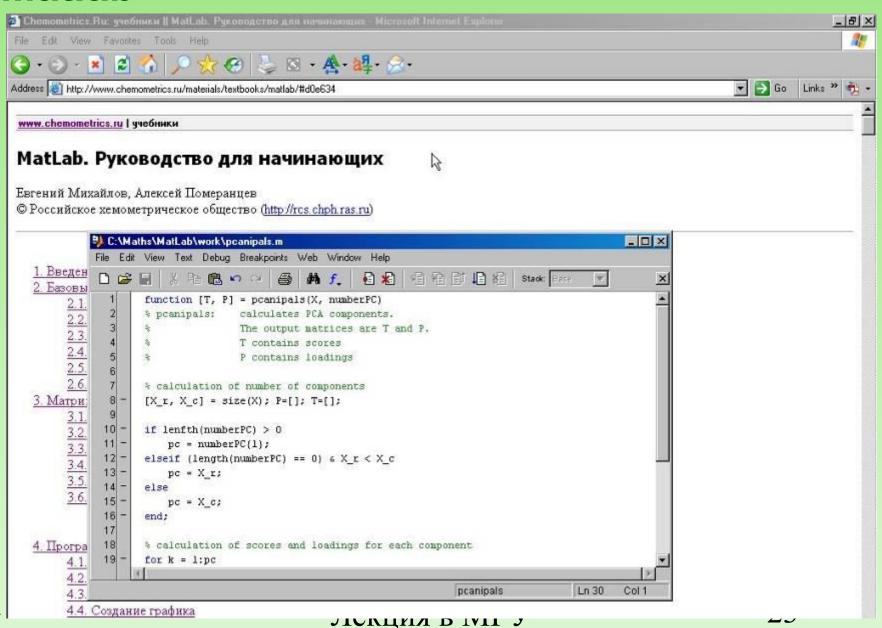


Karl Pearson, 1901

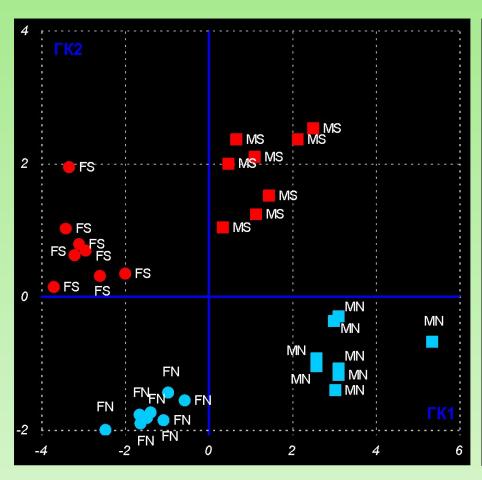
Unscrambler® by CAMO ASA

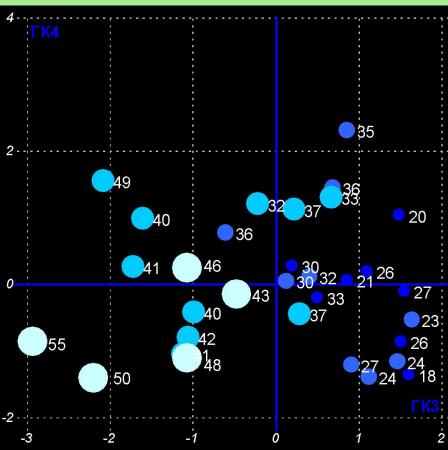


Matlab

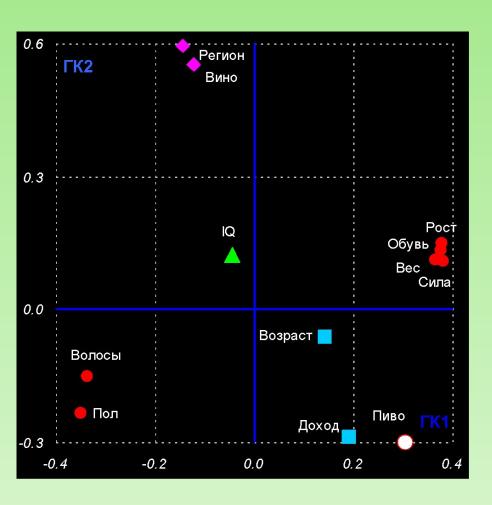


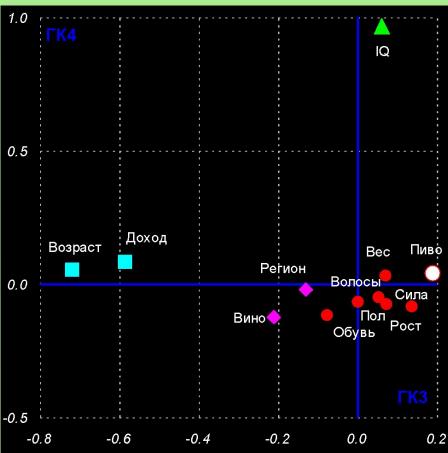
Люди: Графики счетов



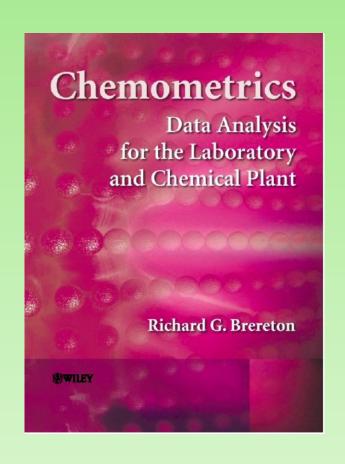


Люди: Графики нагрузок



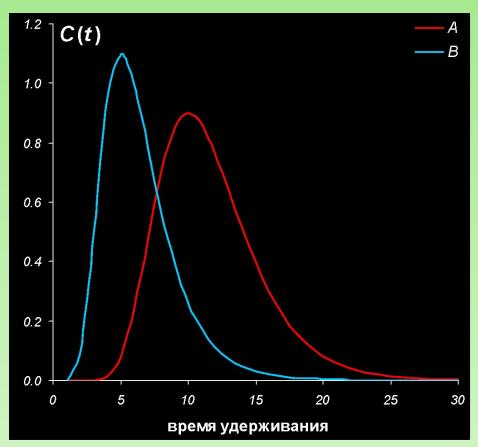


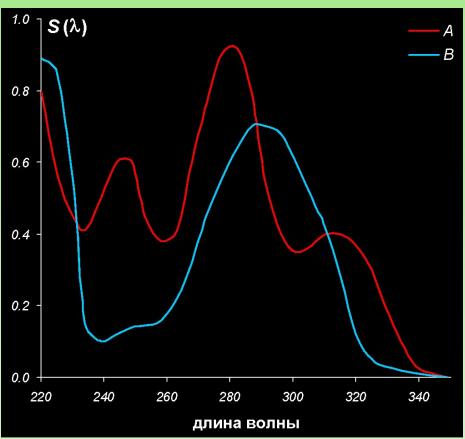
Разделение пиков в ВЭЖХ-ДАМ





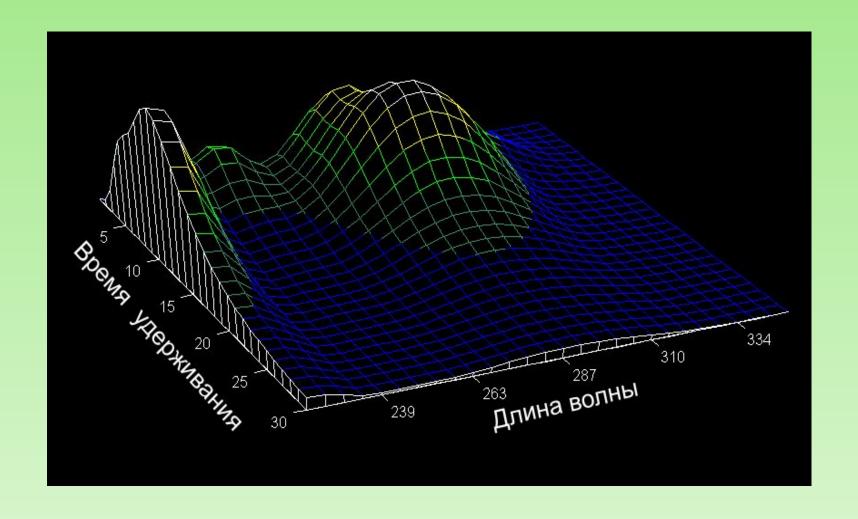
Свойства чистых веществ А и В





X=CS^T+E

Хроматограмма смеси

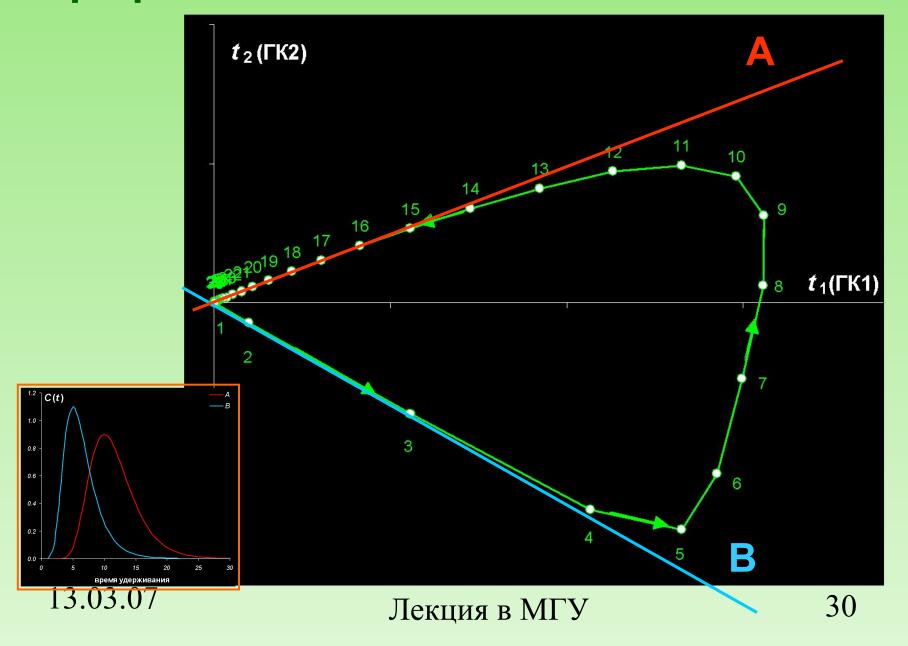


Данные в Unscrambler

The Unscrambler - [HPLC-DAD]												_ D ×						
File Edit View Plot Modify Task Results Window Help														-BX				
		230	234	239	244	249	253	258	263	268	272	277	282	287	291	296	301	306
		3,	4'	5"	6'	7	8.	9'	10°	11'	12'	13'	14'	15"	16'	17"	18'	19'
1	1	0.003	0.003	0.003	0.004	0.004	0.003	0.002	0.003	0.004	0.005	0.006	0.006	0.005	0.004	0.003	0.002	0.01
2	2	0.023	0.021	0.026	0.030	0.029	0.023	0.018	0.021	0.029	0.038	0.045	0.046	0.040	0.030	0.021	0.017	0.0
3	3	0.091	0.085	0.101	0.120	0.117	0.090	0.071	0.084	0.116	0.153	0.178	0.182	0.158	0.120	0.083	0.066	0.01
4	4	0.212	0.198	0.236	0.280	0.271	0.209	0.165	0.194	0.270	0.354	0.413	0.422	0.368	0.279	0.194	0.155	0.11
5	5	0.321	0.296	0.352	0.419	0.405	0.312	0.247	0.291	0.405	0.532	0.621	0.635	0.555	0.422	0.296	0.237	0.2
6	6	0.372	0.323	0.377	0.449	0.435	0.338	0.271	0.321	0.447	0.588	0.689	0.710	0.629	0.490	0.356	0.289	0.21
7	7	0.412	0.299	0.329	0.391	0.382	0.305	0.255	0.310	0.433	0.571	0.676	0.713	0.657	0.546	0.430	0.362	0.3
8	8	0.494	0.270	0.263	0.311	0.311	0.262	0.240	0.304	0.426	0.565	0.682	0.744	0.731	0.662	0.571	0.497	0.4:
9	9	0.586	0.251	0.207	0.243	0.252	0.230	0.233	0.308	0.432	0.576	0.708	0.798	0.824	0.793	0.724	0.643	0.5
10	10	0.628	0.232	0.165	0.193	0.206	0.202	0.222	0.301	0.424	0.568	0.705	0.809	0.860	0.855	0.802	0.719	0.61
11	11	0.606	0.206	0.133	0.155	0.170	0.174	0.201	0.277	0.391	0.524	0.655	0.760	0.819	0.826	0.786	0.707	0.5
12	12	0.544	0.178	0.108	0.126	0.140	0.147	0.174	0.243	0.342	0.460	0.576	0.672	0.729	0.741	0.709	0.638	0.5:
13	13	0.468	0.150	0.089	0.103	0.115	0.123	0.148	0.206	0.291	0.391	0.490	0.573	0.624	0.636	0.610	0.550	0.4:
14	14	0.395	0.125	0.073	0.085	0.095	0.102	0.123	0.173	0.244	0.327	0.411	0.481	0.525	0.535	0.514	0.463	0.3
15	15	0.331	0.105	0.061	0.071	0.080	0.086	0.103	0.144	0.203	0.273	0.343	0.402	0.438	0.447	0.429	0.387	0.3:
16	16	0.277	0.088	0.052	0.060	0.067	0.072	0.086	0.121	0.171	0.229	0.288	0.337	0.367	0.374	0.359	0.324	0.21
17	17	0.234	0.075	0.044	0.051	0.057	0.061	0.073	0.102	0.144	0.193	0.243	0.284	0.309	0.315	0.302	0.272	0.2:
18	18	0.198	0.064	0.038	0.044	0.049	0.052	0.062	0.087	0.122	0.165	0.206	0.241	0.26	0.267	0.256	0.231	0.1!
19	19	0.170	0.055	0.033	0.038	0.042	0.045	0.053	0.074	0.105	0.141	0.177	0.207	0.225	0.229	0.219	0.197	0.11
20	20	0.147	0.048	0.029	0.033	0.037	0.039	0.046	0.064	0.091	0.122	0.153	0.179	0.194	0.197	0.189	0.170	0.1
21	21	0.127	0.042	0.025	0.029	0.033	0.034	0.040	0.056	0.079	0.107	0.133	0.156	0.169	0.171	0.164	0.147	0.11
1																		•
For Help														Value:	0.095	ize: 30 x 2	29 R/W	GU //

13.03.07

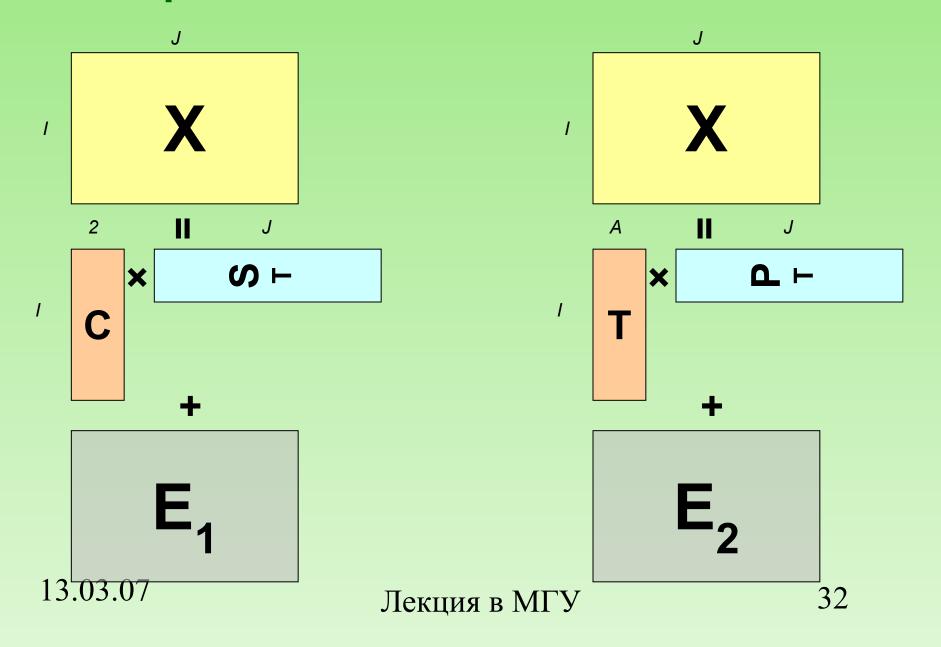
График счетов



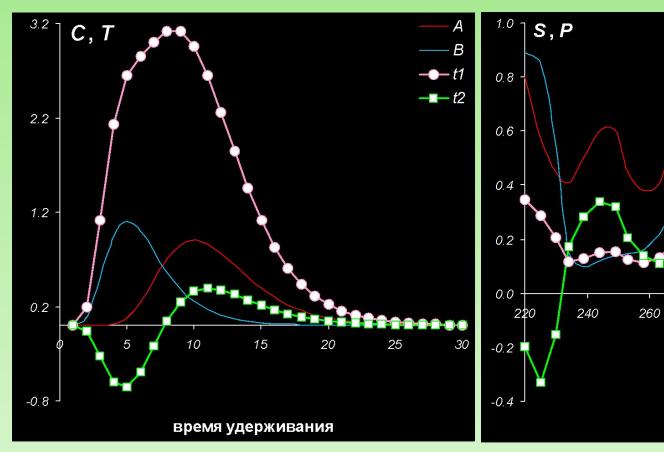
Выводы из графика счетов

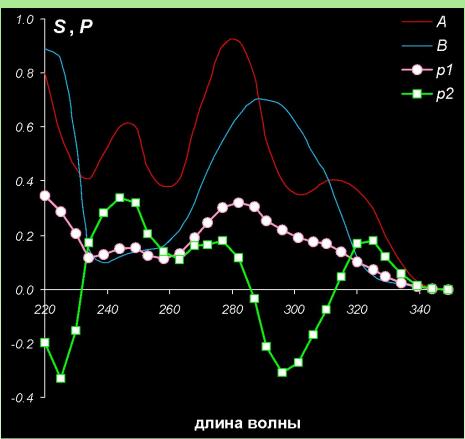
- 1. линейные участки = чистые компоненты
- 2. кривые участки = коэлюция
- 3. ближе к началу = меньше интенсивность
- 4. число поворотов = число чистых компонент

Факторный анализ и анализ ГК

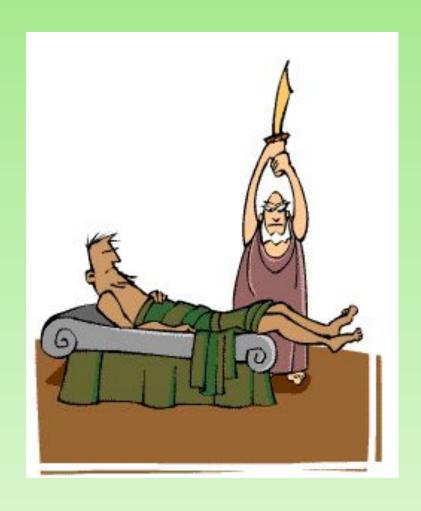


Счета и нагрузки



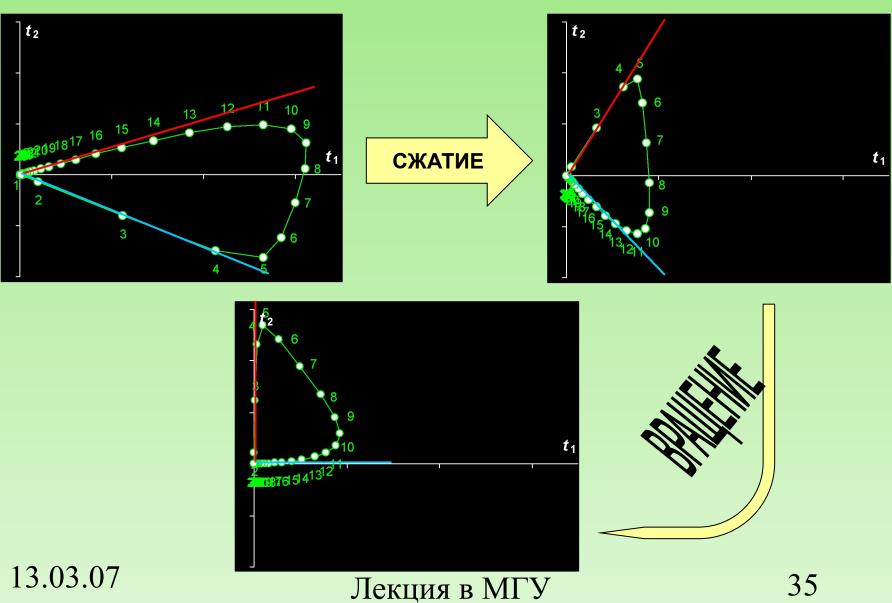


Прокрустово преобразование



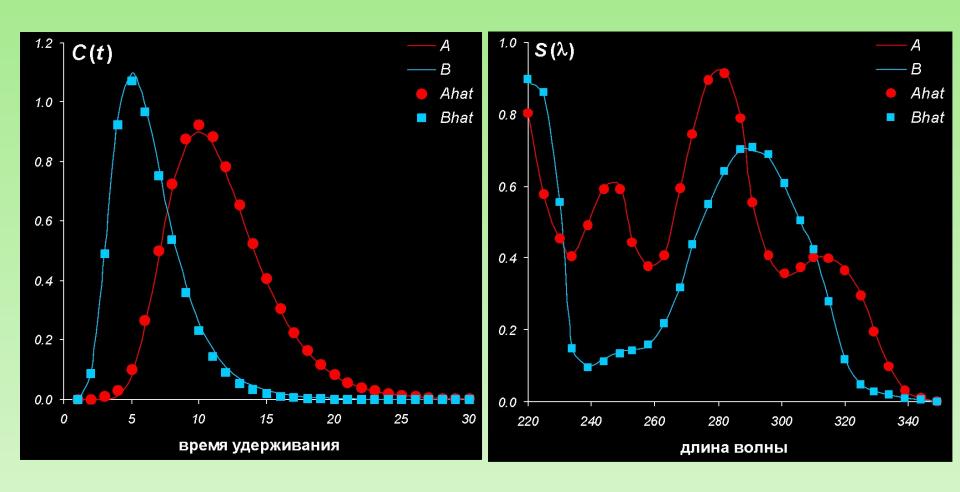
```
X \approx CS^T
X \approx TP^T
I = RR^T = единичная
X \approx T(RR^T)P^T = (TR)(PR)^T
C ≈ TR
                       S≈ PR
R = R_{stratch} \times R_{rotation}
```

Преобразование счетов

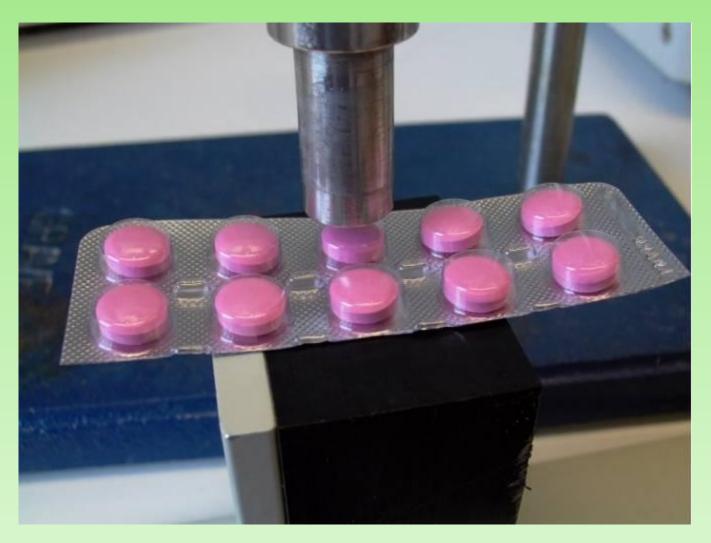


35

Результат Прокрустова преобразования



Классификация поддельных лекарств



Измерения: БИК спектроскопия



Образцы

Подлинники:

11 серий по 5 таблеток

Подделки:

4 серии по 5 таблеток

Всего

15×5= 75 образцов

Обучающий набор

Подлинники:



8 серий по 5 таблеток

Всего:

8×5= 40 образцов

Проверочный набор

Подлинники:



3 серии по 5 таблеток

Подделки:

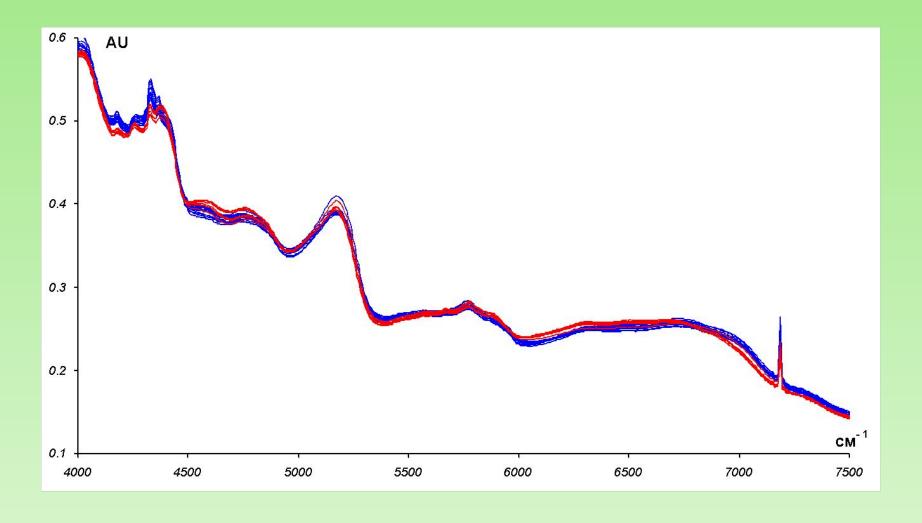


4 серии по 5 таблеток

Всего:

7×5= 35 образцов

БИК спектры мезима



13.03.07

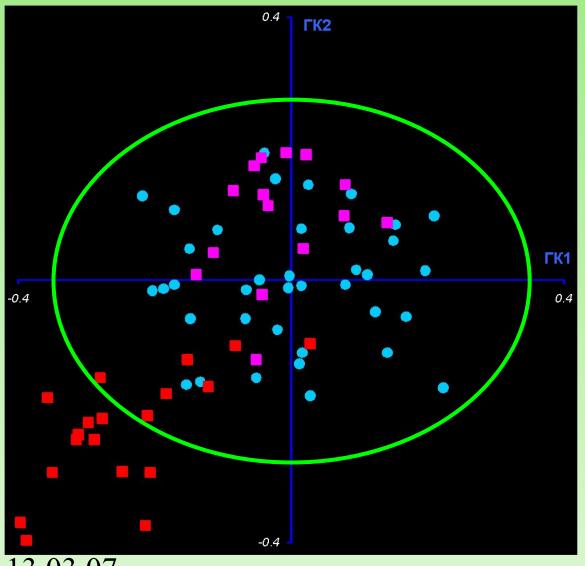
 Настоящие
 Фальшивые

 Лекция в МГУ

Данные: БИК спектры



График МГК счетов



Обучающий набор

Подлинники:



8 серий по 5 таблеток

Всего:

8×5= 40 образцов

Проверочный набор

Подлинники:



3 серии по 5 таблеток

Подделки:

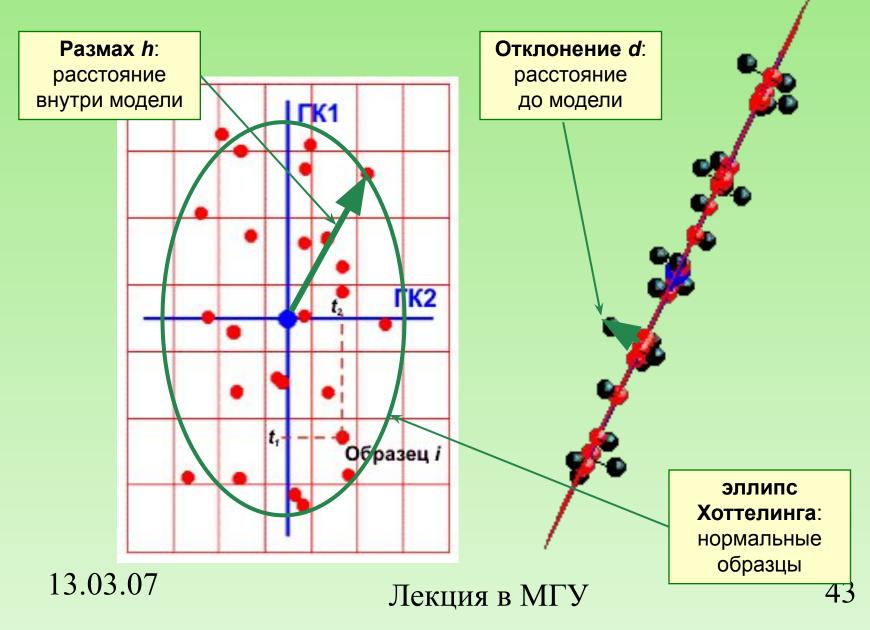


4 серии по 5 таблеток

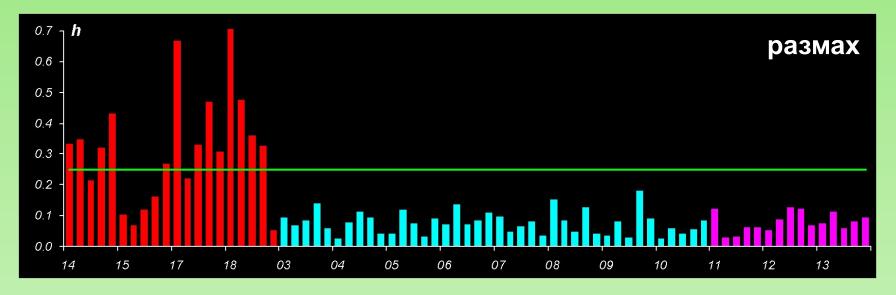
Всего:

7×5= 35 образцов

Размах и отклонение

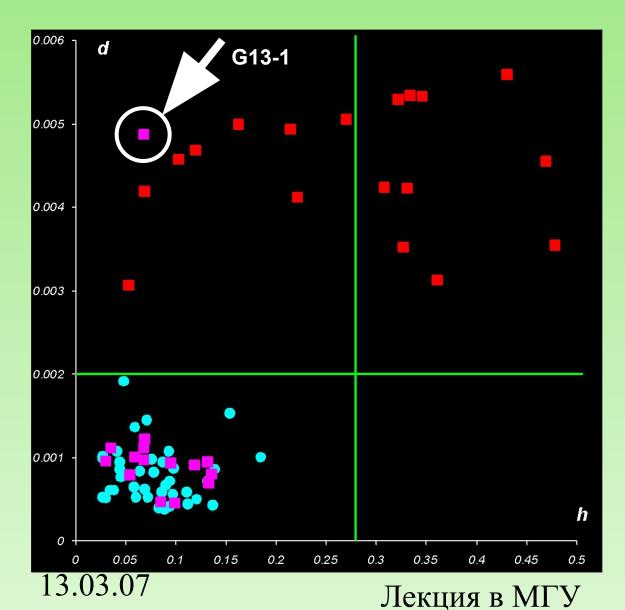


Размах и отклонение для мезима





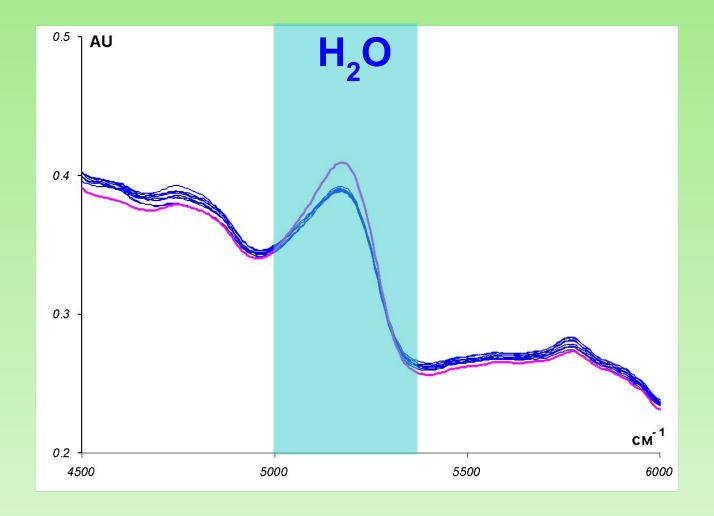
SIMCA



Soft Independent **Modeling of** Class **Analogy** формальное независимое моделирование аналогий классов

45

Выброс G13-1



Хемометрика: два определения

Дедуктивное

Хемометрика - это научная дисциплина, находящаяся на стыке химии и математики, предметом которой являются математические методы исследования химических данных

сайт Российского хемометрического общества

Индуктивное

Хемометрика — это то, что делают хемометрики. сайт Международного хемометрического общества

Хемометрики – это такие люди, которые все время пьют пиво и воруют идеи у математиков

Svante Wold

Что делают хемометрики?

- контролируют производство полупроводников, аспирина и пива;
- исследуют причины разрушения старинных документов;
- проводят допинг контроль спортсменов;
- расшифровывают состав косметики Древнего Египта;
- локализуют месторождение золота в Швеции;
- идентифицируют подозреваемого в убийстве Анны Линд;
- проводят диагностику артрита и рака на ранних стадиях;
- исследуют органические субстанции в кометном веществе;
- исследуют кормовой рацион на свинооткормочных комплексах;
- проверяют, как рацион питания влияет на умственные способности;
- определяют фальшивые лекарства;
- и еще многое, многое другое

А если серьезно

- Хемометрика имеет дело с данными (зачас-тую с очень большими), поэтому хемометрика это подраздел информатики (Data mining)
- Данные, которые исследует хемометрика по большей части происходят из **химии**, поэтому хемометрика это подраздел химии (Analytical chemistry)
- Методы, которые использует хемометрика ориентированы на формальное моделирование (Soft modeling)

Два «не» и три «да»

- 1. Хемометрика ≠ химическая метрология
- 2. Хемометрика ≠ статистика в химии

Хемометрика решает следующие задачи в области химии:

- (1) как получить химически важную информацию из химических данных,
- (2) как организовать и представить эту информацию,
- (3) как получить данные, содержащие такую информацию.

Что можно почитать ...

Хемометрика: достижения и перспективы О.Е.Родионова, А.Л.Помераниев Инстатут химической физики им. Н.Н.Семенова Российской академии наук 119991 Москва, ул. Косыгина, 4, факс (495)939-7483 Рассмотрены основные хемометрические методы в модели, непользуемые для решения задач качественного и количественного зислита, а тлеже дли выглитического ком тиоля технологических процессов. Покладым достижения в области хемометрики за последние 20 лет. Обеуждены тенденции и певенективы ее патинтии. Отлавление Выдолен

И. Данные и модели, осносъзуваные в запячествої зналиче

И. Методы вочественного знализь. Исследования, изволяфикац

ГУ. Методы комечестваного внализь. Градуировочные водделя

У. Подготовета данные и обработих сигнализя

VI. Заклагачана



Ким Эсбенсен **СИПАНА МНОГОМЕРНЫХ** ДАННЫХ



Хемометрика у нас..



18 man 2005 r

• 970 Таков камонотрена · WHY SHIP SAFETY • парки Общество

І. Введение

1. Истории хемометрики и ее место в системе знавий Со премени опубликования перевода на русский жых единетненной (до веданиего премени) жинги по хемометрике " пре

- организационнай компис
- · remain Version PACO
- · DECECHARANAM CHARACTERS
- принциплиний дання
- · Britanness PRO
- appear FXQ
- култы по хомометрине
- · Chemometrica 90
- внализинская жимия в России

...у ник...

- · CAN VOHETS TAD THROMETORY
- · INHOGENIAM DE LEMORSTERE
- партика и национальные общества • Курнальско хемаметрике
- программите обеспечения

... в топерь вместе!

- · Drustparretrics project
- DM40.5
- · WSG-I
- · WS0-2 W60-3
- WSC-4

Российское Хемометрическое Общество



изиболее важное напривление в земометрике. В последнее премя оно очень быстро и плодотнорно развинеется, при этом зимикам-виалитикам предложены не топью неные

методы обработки данных, но и новые полходы в постановке

томен во стыте пимов ^в в мотемотите, в кое это восто быто с

параже на ставое замния "и матементика, и всек иго часто бывает се пограничанным доставитамима, до сок поер не моет общего с пограничанного опредставия. Наяболяет популярное определение приматамит д. Имостаруу, "который ставта, что име приматамит д. Имостаруу, "который ставта, что замниметрико — это замническая дисприлины, в которой при-метняют метлема ическаме, ставтелические и другие метолом метнямит метольным ическаме, ставтелические и другие метолом метнямит метольным ическаме, ставтелические и другие метолом другие метолом

Наделево по Nove Science Райзіння выпустко об орган матринов Треть ві пастей пастей пастей п assessments (WOCA) tragment sense Progress in Chemometrics Research.

Попиционентел и подпровънност вишто и таконить ет может на сайте правтив ства

Посмотреть огливление

Прочитать предисление

Bristol

20 апреда 2005 г.

19 априля 2005 г. в раменя выстания Аподуйльборо присши однодновная конференция. "Хемпьютрова и контроль принтодеть опыск процестов". Протограции профессора R. Восебох (Индерсору of Bristol, URC "PAT – Тонгаватия

(бегоды) енеквия прицегосо" и другия дикладовиля возная полнопреть на сейте-DODGE FARRIES

Приграммы конференции

2 марта 2005 г

Свидумиря поонай политиры и паказа по полимеряю "Скарывнова и метцал вилас прилговиров ст дискост (WSC-5) пробарт 19-23 фиграля 2001 г. в Сченую

Тимически конфиционар АТОКОРС, переработки отпоров и паката кануженский среды.







+ WSC-4 Chemometrics, ku

+ 0.came + Состиваниемся

S and

(i) 100

+ Соменниорновы

• Что такое оневентрика? + Fepmentones

-Хожнетр-ка

+ FROZEI

* + Crames

A + F000

i + (boxeriaze

н Канференция

Orambbametrica

+ 0.000ecte

* DM 0.5

+ WSC-1

+ 9502

· WSC-2



Bat/ex

Chemometrics.Ru :: хемометрика в России

Новости

2005-05-26 // Изменения

В редделе "статьи" дополнен и исправлен ранее опубликованный доклад Карпунина О.Н. на четвертом махиучнародном эммнем симпоэкоме по хемометрике (МБС-4).

2005-05-04 // Новые презентации на сайте

19 апреля 2005 г. в рамках выставки AnalyticaExpo прошла <u>однодневная конференция</u> "женометрика и контроль производственных процессов", с участнем профессора Бристольского университета Ричарда Бреретона, Презентации докладов и лекций, прочитанных на конференции вы можете найти в соответствующем разделе по ключевой фразе "апаlyticaexpo" или перейдя по этой ссылке,

2005-04-22 // Презентации с WSC-IV

На сайте выпожена большая часть презентаций докладов и ленций, прочитанных участниками четвертого международного симпозиума по жемометрике WSC-IV. Вывести весь относк доступных на данное время презентаций можно сделав поиск по ключевому слову "WSC-IV" в резделе <u>Презентации</u> хибо использую <u>эту ссылку</u>.

2005-04-07 // Пополнение в разделе "Статьи"

В разделе "статьи" выпожен доклад, который был сделан д.х.н., заведующим лабораторией института кимической физики РАН Карпухиным Олегом Никифоровичем на закрытии четвертого международного симпозиума по хемпентрике УСС-4. Е статье поднимаются очень интересные вогросы о роли хемометрики в современной науке и возможных вармантах ее дальнейшего развитил. Очень интереснал точка оренил, поонакомится с которой будет полеоно как начинающим ученым, так и тем, кого уже можно назвать опытным исследпентелем.

Биновойшие событив

27.06.2005

BC's Znd Aroual Process Analytical Technologies for

Droppe exemples roudepeaux and отещевлистов из провышленности PACEGIX HARRISTERIOS. митерисующихся и работающих из MICHIGOROPOUS ET STRUTH PROCESSORS контрей в презыкстве - Рессиях anietical Technologies (PAT):

Глучайная ссывка

Homegage of Chemometrics - conws classic transplagment delitocoerane a ресурсов по хемометрика. Содирант очень вино то реансобрасной информации навасти, персоналии, конференции статын н.т.д. Создалель и редастор -Johan Trysa

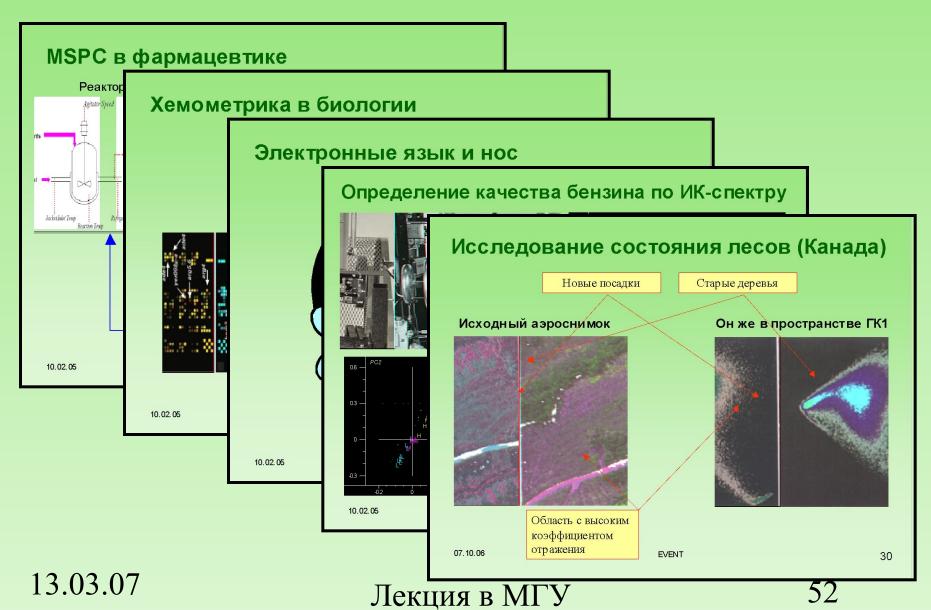
Новое на сервере

В унодело Канференции поряжнось информация с второй ожи тодной Process Analytical Technologies (PAT) p Suppoper

Розден Термены положения MICCOTTACIONE HORIZON MATERIALISMOS

В раздала Конференции полектась жеформицио о серинарофия конференции "Стотистики в замени

Что осталось на потом ...



Спасибо за внимание!

