

УДК: 519.237.07

Конфирматорная факторная модель артериальной гипертензии

В. А. Шовин

Омский филиал Учреждения Российской академии наук
Института математики им. С. Л. Соболева Сибирского отделения РАН,
Россия, 644099, г. Омск, ул. Певцова, д. 13

E-mail: v.shovin@mail.ru

*Получено 29 апреля 2012 г.,
после доработки 19 октября 2012 г.*

Предлагается новая методика построения ортогональной факторной модели на основе метода корреляционных плеяд и конфирматорного факторного анализа. Предложен новый алгоритм конфирматорного факторного анализа. На основе оригинальной методики построена факторная модель артериальной гипертензии первой стадии. Проведен анализ корреляционных зависимостей и показателей артериальной гипертензии.

Ключевые слова: артериальная гипертензия, корреляционные плеяды, конфирматорный факторный анализ

Confirmatory factor model of hypertension

V. A. Shovin

Omsk Branch of the Institution of the Russian Academy of Sciences Institute of Mathematics. S. Siberian Branch of RAS, Russia, 644099, Omsk, Pevtsov, 13

Abstract. — A new method of constructing orthogonal factor model based on the method of correlation pleiades and confirmatory factor analysis. A new algorithm for confirmatory factor analysis. Based on an original method built factor model of hypertension the first stage. The analysis of correlations and indices of arterial hypertension.

Keywords: hypertension, pleiad of correlation, confirmatory factor analysis

Citation: *Computer Research and Modeling*, 2012, vol. 4, no. 4, pp. 885–894 (Russian).

Введение

В данной работе исследуется феномен артериальной гипертензии начальной стадии, причиной которой не является патология органов. Главным диагностическим признаком артериальной гипертензии (АГ) является повышенное артериальное давление. Артериальная гипертензия является одной из наиболее значимых медико-социальных проблем Российской Федерации [Чазова, 2008].

В качестве исходных данных научно-исследовательской работы использовался набор показателей, характеризующих артериальную гипертензию I стадии I степени. Данный набор показателей был получен в ходе клинического обследования лиц с данной патологией.

Основной целью данной научно-исследовательской работы является выделение групп взаимосвязанных параметров, которые определяют главные направления лечения. Для достижения поставленной цели предлагается использовать новую методику последовательного применения метода корреляционных плеяд и конфирматорного факторного анализа. На базе корреляционных плеяд осуществляется разделение параметров на отдельные группы. С помощью конфирматорного факторного анализа осуществляется поиск весовых коэффициентов факторов, представленных этими группами переменных.

Метод корреляционных плеяд реализуется на базе алгоритма Прима: построение максимального корреляционного пути для матрицы корреляций исследуемых признаков.

Предлагается новый вариант конфирматорного факторного анализа. В качестве критерия конфирматорного факторного анализа предложен функционал — сумма квадратов невязок между исходной и восстановленной корреляционных матриц. Для поиска минимального значения этого функционала с ограничениями на элементы факторной структуры был выбран метод случайного поиска. Гипотеза о виде факторной структуры представлена моделью корреляционных плеяд.

Метод корреляционных плеяд

Одним из методов разбиения признаков на группы является метод корреляционных плеяд. Плеяда — это группа переменных, корреляционная связь между которыми (внутриплеядная связь) достаточно велика. Связь между параметрами из разных групп (межплеядная) — мала.

Исходными данными для метода корреляционных плеяд является матрица корреляций между переменными. Коэффициенты корреляций рассматриваются по модулю.

Выделение плеяд можно осуществить способом максимального корреляционного пути.

В этом случае все m признаков располагаются таким образом, чтобы сумма $(m-1)$ связей между ними являлась наибольшей [Айвазян, 1989]. Эту задачу предлагается решать с помощью алгоритма Прима.

Алгоритм Прима

Данный алгоритм позволяет соединить m вершин графа $(m-1)$ связями, так чтобы можно было из любой вершины добраться до любой другой. При этом требуется максимизировать сумму связей между вершинами.

Пусть $U = \emptyset$ — пустое множество,

V — множество вершин графа,

(v, u) — ребро графа,

$P = \emptyset$ — пустое множество.

1. Выбрать (v, v') — ребро наибольшей стоимости. Добавить ребро (v, v') в P .
2. Добавить v и v' в U .

3. Пока $V \setminus U$ не пусто, выбрать ребро (u, v) наибольшей стоимости, где $u \in U$, $v \in V \setminus U$. Добавить v к U . Добавить ребро (u, v) в P .

4. Далее повторять пункт 3) до тех пор, пока $V \setminus U$ не пусто.

В результате получаем множество ребер $(u, v) \in P$.

Для исследуемой задачи вершины графа представлены признаками, а связи между вершинами коэффициентами корреляций. После применения алгоритма Прима получим $(m-1)$ коэффициентов корреляции связывающих m признаков, образующих остовное дерево, в котором все признаки расположены таким образом, что связи между ними являются наибольшими. Это и есть максимальный корреляционный путь.

Если задать порог для абсолютного значения коэффициента корреляции в остовном дереве, то дерево распадется на плеяды.

Конфирматорный факторный анализ

Конфирматорный факторный анализ это особый случай факторного анализа. При построении конфирматорной факторной модели вводятся дополнительные ограничения на значения элементов факторной структуры, что позволяет доказать или опровергнуть гипотезу о виде факторной структуры исследуемого объекта на выборочном эмпирическом материале.

Факторный анализ основывается на следующей линейной модели, связывающей исходные показатели Z и факторы P :

$$Z = AP, \tag{1}$$

где $Z \leftrightarrow z_{ij}$ — матрица размерности $m \times n$ значений m параметров у n объектов,

$P \leftrightarrow p_{ij}$ — матрица размерности $g \times n$ значений g факторов у n объектов,

$A \leftrightarrow a_{ij}$ — матрица факторной структуры размерности $m \times g$ весовых коэффициентов.

В ортогональной факторной модели предполагается, что факторы ортогональны, т. е. $PP^T = E$ — единичная матрица.

Следовательно, матрица $R = ZZ^T$ корреляций переменных определяется согласно теореме Терстоуна [Иберла, 1980]:

$$R = AA^T. \tag{2}$$

Методика конфирматорного факторного анализа состоит в том, что выдвигается гипотеза о виде факторной структуры, т. е. о том, какой фактор имеет нагрузку и от каких переменных. Решением конфирматорного факторного анализа является такая факторная структура, которая наиболее полно согласуется с наблюдениями при том, что общий вид ее строго определен [Мьюллер, Клекка, 1989].

В качестве критерия, определяющего степень правдоподобия факторной модели эмпирическим данным, выбирается критерий минимальных остатков, т. е. сумма квадратов невязок между элементами эмпирической и восстановленной корреляционных матриц.

Восстановленная корреляционная матрица R_v вычисляется исходя из теоремы Терстоуна.

Предлагаемый критерий конфирматорного анализа имеет вид

$$f = \sum_{i=1}^{m-1} \sum_{j=i+1}^m (r_{ij} - \sum_{k=1}^g a_{ik} a_{jk})^2 \rightarrow \min, \tag{3}$$

r_{ij} — коэффициенты корреляций между переменными,

a_{ij} — элементы факторной структуры,

m — число переменных,

g — число факторов.

Задача конфирматорного факторного анализа теперь сводится к задаче нелинейной оптимизации функции $f(a_{ij}; i=1\dots m, j=1\dots g)$ с дополнительными ограничениями:

$$1) -1 \leq a_{ij} \leq 1; i=1\dots m, j=1\dots g;$$

$$2) \sum_{k=1}^g (a_{ik})^2 \leq 1; i=1\dots m;$$

$$3) s_{ij} \leq a_{ij} \leq t_{ij}; i=1\dots m, j=1\dots g;$$

Ограничения 1) и 2) — это общие ограничения факторного анализа соответственно о значениях факторной структуры и величин общностей переменных.

Ограничение 3) представляет собой гипотезу о виде факторной структуры.

Нахождение глобального минимума функции f предлагается проводить с помощью метода случайного поиска. В качестве случайных точек предлагается использовать псевдослучайные последовательности [Медынский, Антоний, 2003].

Генерация псевдослучайных последовательностей

Для генерации случайных последовательностей предлагается использовать генератор случайных чисел $I(n)$, основанный на рекуррентной последовательности (линейный конгруэнтный метод) [Кнут, 2000]:

$$I(n+1) = (a * I(n) + c) \pmod{m}. \quad (4)$$

Число a называется мультипликатором, число c — инкрементом, а число m — модулем.

\pmod — получение остатка от деления целочисленных операндов.

Чтобы получить диапазон случайных чисел $y \in [0,1]$, используют преобразование

$$y = I/m.$$

В качестве a, c, m можно использовать следующие значения:

$$a = 4096,$$

$$c = 150889,$$

$$m = 714025.$$

Алгоритм случайного поиска

Метод случайного поиска заключается в последовательной генерации набора равномерно распределенных независимых случайных величин. Каждая случайная величина соответствует отдельной неизвестной переменной в оптимизируемом функционале. Для каждого набора случайных величин значение функционала сравнивается с наилучшим, найденным на предыдущих итерациях. Если новое значение функционала оказывается лучше, то оно принимается в качестве наилучшего. Генерация набора случайных величин осуществляется в локальной области с центром в наилучшей точке, полученной на предыдущей итерации. В ходе выполнения итераций осуществляется также сужение локальной области. Процесс повторяется до тех пор, пока не будут выполнены все итерации. Далее описан алгоритм случайного поиска, адаптированный для задачи конфирматорного факторного анализа.

Переход от матричного представления переменных функции f к векторному осуществляется с помощью преобразования индексов:

$$a_{ij} = x_{(i-1)g+j}.$$

1. Задаем область изменения переменных $x_i \in [a_i, b_i]$.
При этом интервал $[a_i, b_i] \subset [s_i, t_i]$.
2. Задаем Δ , например, равным 0.1.
3. Создаем набор векторов $\bar{x} = (x_1, x_2, \dots)$ псевдослучайных чисел $\in [0, 1]$
4. Переводим в диапазон $[a_i, b_i]$ с помощью формулы

$$x_i = a_i + x_i \cdot (b_i - a_i).$$

5. Среди набора различных векторов псевдослучайных чисел ищем тот x_{\min} , что доставляет минимум функции f и удовлетворяет дополнительным условиям 1 и 2 факторного анализа.
6. Если на данной итерации не улучшен критерий f , то уменьшают Δ .
7. Создаем новые интервалы $[a_i, b_i]$ по формулам

$$a_i = x_{\min i} - \Delta,$$

$$b_i = x_{\min i} + \Delta.$$

При этом дополнительно ограничиваем новый интервал условием $[a_i, b_i] \subset [s_i, t_i]$.

8. Переходим к пункту 3.

Примечание:

Последовательность уменьшающихся Δ должна образовывать расходящийся ряд [Миркес, 2002]. Поэтому в качестве такой последовательности может служить ряд $\frac{1}{10+n}$, где n — число обращений к пункту 6).

Алгоритм конфирматорного факторного анализа

Алгоритм конфирматорного факторного анализа представлен следующими пунктами.

1. Исходными данными является матрица R корреляций между m переменными.
2. Фиксируется число факторов g .
3. Формируются ограничения 3) на значения элементов факторной структуры.
4. В качестве критерия выбирается функция (3).
5. Производится минимизация критерия с дополнительными ограничениями с использованием алгоритма случайного поиска и генератора псевдослучайных последовательностей.

Методика построения факторной модели

В данном разделе излагается оригинальная методика построения факторной модели на базе модели корреляционных плеяд и конфирматорного факторного анализа. Данная методика не имеет отечественных и зарубежных аналогов. Суть методики заключается в следующем:

- 1) вычисляется матрица R корреляций признаков;
- 2) строится корреляционный дендрит по принципу максимального корреляционного пути;
- 3) полученный дендрит разбивается на плеяды;
- 4) выбирается пороговый уровень корреляции 0.7;
- 5) формируется гипотеза для конфирматорного факторного анализа на базе информации о наполнении переменными выделенных групп;
- 6) в рамках сформированной гипотезы осуществляется поиск факторной структуры с использованием оригинального алгоритма конфирматорного факторного анализа.

Замечание. Каждой выделенной группе сопоставляется отдельный фактор. Если некоторая переменная i входит в фактор j , значит соответствующий коэффициент факторной структуры

по абсолютному значению предполагается принадлежащим интервалу $|a_{ij}| \in [0.7, 1]$, в противном случае — $|a_{ij}| \in [0, 0.7]$.

Исходные медицинские данные

В качестве исходных параметров были взяты 35 биофизических показателя для 130 лиц с артериальной гипертензией начальной стадии. Для этих показателей были вычислены коэффициенты корреляции Пирсона. Из исходного набора параметров были исключены переменные, корреляционная связь между которыми была меньше 0.7. Переменные, не образующие корреляционную связь с другими переменными, фактически представляют факторы. Поэтому не имеет смысла оставлять такие переменные для факторного исследования. Параметр оставляется, если он имеет коэффициент корреляции больше 0.7 хотя бы с одним другим параметром.

В результате были выделены следующие параметры объектов:

- 1) вес,
- 2) индекс массы тела (ИМТ),
- 3) частота дыхания (ЧД),
- 4) сегментоядерные нейтрофилы (С),
- 5) лимфоциты (Л),
- 6) конечно-систолический размер левого желудочка (КСР),
- 7) конечно-систолический объем левого желудочка (КСО),
- 8) конечно-диастолический размер левого желудочка (КДР),
- 9) конечно-диастолический объем левого желудочка (КДО),
- 10) ударный объем (УО),
- 11) минутный объем сердца (МОС),
- 12) общее периферическое сосудистое сопротивление (ОПСС),
- 13) индекс Хильдебрандта (ИХ),
- 14) фракция выброса левого желудочка (ФВ),
- 15) фракция укорочения левого желудочка (ФУ).

Для этих параметров была сформирована индивидуальная матрица корреляций (табл. 1).

Таблица 1. Матрица коэффициентов корреляций между признаками

	Вес	ИМТ	ЧД	С	Л	КСР	КСО	КДР	КДО	УО	МОС	ОПСС	ИХ	ФВ	ФУ
Вес	1.00														
ИМТ	0.72	1.00													
ЧД	0.30	0.07	1.00												
С	-0.19	-0.13	-0.05	1.00											
Л	0.12	0.06	-0.03	-0.92	1.00										
КСР	0.35	0.05	0.21	0.18	-0.23	1.00									
КСО	0.29	0.10	0.36	0.28	-0.31	0.71	1.00								
КДР	0.36	0.25	0.12	0.18	-0.27	0.73	0.78	1.00							
КДО	0.38	0.13	0.28	0.23	-0.26	0.83	0.93	0.87	1.00						
УО	0.42	0.14	0.18	0.17	-0.18	0.84	0.77	0.85	0.94	1.00					
МОС	0.24	-0.05	0.30	0.28	-0.26	0.77	0.73	0.62	0.86	0.88	1.00				
ОПСС	-0.23	-0.07	-0.00	-0.29	0.20	-0.63	-0.60	-0.61	-0.72	-0.75	-0.66	1.00			
ИХ	-0.33	-0.06	-0.81	0.11	-0.01	-0.14	-0.26	-0.14	-0.22	-0.18	-0.16	0.04	1.00		
ФВ	0.07	0.05	-0.19	-0.19	0.29	-0.35	-0.40	-0.18	-0.16	0.04	0.01	0.02	0.05	1.00	
ФУ	-0.16	-0.12	-0.33	0.04	0.124	-0.42	-0.41	-0.26	-0.24	-0.07	-0.01	0.06	0.11	0.80	1.00

Результаты работы конфирматорной факторной модели и метода корреляционных плеяд

По данной корреляционной матрице на уровне корреляции 0.7 были выделены плеяды (рис. 1), образованные следующими пятью группами взаимосвязанных параметров:

- 1) конечно-систолический размер левого желудочка,
 конечно-систолический объем левого желудочка,
 конечно-диастолический размер левого желудочка,
 конечно-диастолический объем левого желудочка,
 ударный объем,
 минутный объем сердца,
 общее периферическое сосудистое сопротивление;
- 2) фракция выброса левого желудочка,
 фракция укорочения левого желудочка;
- 3) вес,
 индекс массы тела;
- 4) частота дыхания,
 индекс Хильдебранта;
- 5) сегменты,
 лимфоциты.

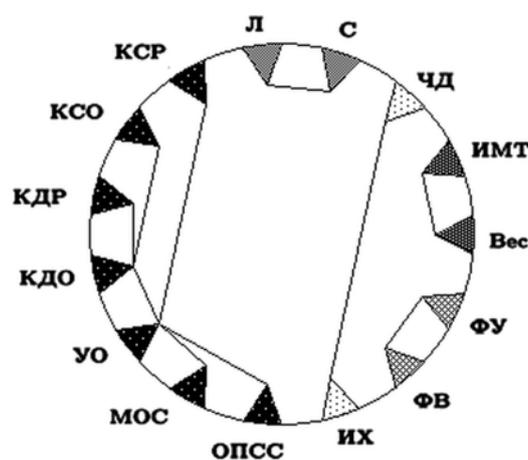


Рис. 1. Корреляционные плеяды на уровне корреляции 0.7 для исследуемых переменных. Треугольниками обозначены исследуемые переменные, а линиями — связи из максимального корреляционного пути, превышающие уровень корреляции 0.7

Отдельные группы переменных были интерпретированы как различные факторы. С помощью информации о принадлежности переменных к различным группам была сформирована гипотезы о виде факторной структуры.

С помощью оригинального алгоритма конфирматорного факторного анализа было найдена ортогональная факторная структура (табл. 2). В рамках сформированной гипотезы, основанной на модели корреляционных плеяд, был достигнут критерий $f = 0.0144490901$.

Факторы, образованные выделенными группами переменных проинтерпретированы следующим образом.

1. Главный фактор можно интерпретировать как гемодинамический фактор, включающий параметры, описывающие центральную и периферическую гемодинамику. Переменные УО, МОС, ОПСС определяют уровень артериального давления. В норме изменениям минутного объема циркуляции должна соответствовать адекватная по величине и направлению реакция прекапиллярного русла, которая бы нивелировала эти изменения и сохраняла среднее давление на нормальном уровне. Например, если МО снижен, то артериолы должны сузиться. Если МО увеличен, то артериолы должны расшириться. Нарушения взаимосвязи этих показателей лежат в основе изменений уровня АД. [Фофанов, 1977] Вместе с тем изменение уровня артериального давления взаимосвязано с модуляцией сердца, за которую отвечают параметры КСР, КСО, КДР, КДО.

Таблица 2. Факторная структура конфирматорного факторного анализа

	F_1	F_2	F_3	F_4	F_5
Вес	0.294	0.364	-0.159	-0.144	0.700
ИМТ	0.255	0.392	-0.207	-0.170	0.713
ЧД	0.136	0.379	-0.313	0.729	0.046
С	0.065	-0.887	-0.406	0.138	0.151
Л	-0.052	0.738	0.376	-0.120	-0.127
КСР	0.785	0.231	-0.394	-0.208	-0.167
КСО	0.779	0.215	-0.455	-0.196	-0.201
КДР	0.957	0.135	-0.081	-0.094	-0.144
КДО	0.960	0.142	-0.139	-0.089	-0.145
УО	0.968	0.040	0.194	0.043	-0.060
МОС	0.898	0.025	0.194	0.027	-0.128
ОПСС	-0.700	-0.016	-0.198	-0.002	0.014
ИХ	-0.056	-0.324	0.309	-0.700	-0.042
ФВ	-0.106	-0.256	0.766	0.382	0.201
ФУ	-0.053	-0.261	0.793	0.226	0.151

2. Фактор составленный из параметров *фракция выброса левого желудочка* и *фракция укорочения ЛЖ* можно считать важным для непосредственной оценки контрактильной (сократительной, нагнетательной) функции левого желудочка. Этот фактор определяет объемную ресурсоемкость ЛЖ. Он показывает, насколько использованы объемные резервы самого сердца для поддержания уровня артериального давления.

3. Фактор, отвечающий за соответствие массы и роста.

4. Фактор, характеризующий уровень слаженности работы сердца и легких, определяется через частоту дыхания и индекс Хильдебранта.

5. Иммунологический фактор, который может отражать психосоматическое состояние индивида, поскольку этот фактор активируется в стрессовых состояниях. Основной вклад в формирование этого фактора вносят сегментоядерные нейтрофилы и лимфоциты.

Сравнение конфирматорной факторной модели и модели, полученной методом главных компонент с варимакс-вращением

Применяя к матрице коэффициентов корреляций (табл. 1) метод главных компонент с варимакс-вращением, получим следующую матрицу факторной структуры (табл. 3):

Таблица 3. Факторная структура варимакс

	F_1	F_2	F_3	F_4	F_5
Вес	0.152	0.073	-0.084	0.074	0.856
ИМТ	0.109	0.084	-0.137	0.074	0.857
ЧД	0.081	0.046	-0.068	0.850	0.084
С	-0.027	-0.916	-0.013	-0.046	-0.094
Л	0.028	0.918	0.034	0.018	0.060
КСР	0.752	0.018	-0.556	0.056	0.145
КСО	0.741	-0.014	-0.602	0.079	0.117
КДР	0.953	0.027	-0.228	0.037	0.101
КДО	0.945	0.007	-0.276	0.062	0.111
УО	0.966	0.019	0.096	0.043	0.088
МОС	0.933	0.030	0.092	0.016	0.010

	F_1	F_2	F_3	F_4	F_5
ОПСС	-0.739	-0.026	-0.150	0.022	-0.088
ИХ	-0.002	-0.014	0.056	-0.857	-0.052
ФВ	-0.056	0.008	0.896	0.008	-0.093
ФУ	0.015	0.045	0.873	-0.135	-0.102

Число факторов определялось по критерию Гутмана (рис. 2). Критерий Гутмана определяет нижнюю границу числа g факторов, подлежащих выделению: если у корреляционной матрицы число собственных значений, больших единицы, равно s , то имеет место соотношение $g \geq s$. Критерий (3) для данной факторной структуры имеет несколько худшее значение $f = 0.0196$ по сравнению со значениями для конфирматорной факторной структуры $f = 0.0144$. Интерпертабельность факторной структуры представляется несколько лучшей. В целом алгоритм конфирматорного факторного анализа с использованием гипотезы корреляционных плеяд является альтернативной методикой построения факторной модели.

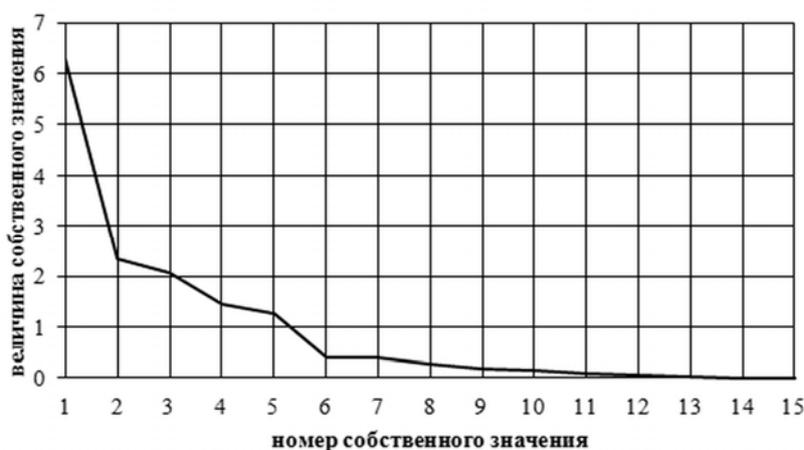


Рис. 2. Число факторов по критерию Гутмана равно 5.

Заключение

Предложена альтернативная методика построения ортогональной факторной модели на основе метода корреляционных плеяд и конфирматорного факторного анализа. Разработан новый алгоритм конфирматорного факторного анализа с использованием критерия минимальных остатков и метода случайного поиска для решения нелинейной оптимизационной задачи. На основе оригинальной методики построена факторная модель артериальной гипертензии первой стадии.

Выделены пять факторов, которые позволяют указать группу параметров, на которую нужно воздействовать, чтобы получить максимальный эффект от лечения. Например, для стабилизации уровня артериального давления следует воздействовать на всю группу признаков, описывающих гемодинамический фактор. При этом следует учитывать ремоделирование сердца (структурно-геометрическое состояние) при формировании патофизиологических взаимоотношений в системе кровообращения у пациентов с гипертонической болезнью. Поскольку ожирение является одним из факторов риска, то снижение веса позволит нормализовать фактор, отвечающий за соответствие массы и роста. Исключение стрессовых ситуаций пациентом позволит улучшить показатели, формирующие иммунологический фактор, а также нормализовать фактор, характеризующий уровень слаженности работы сердца и легких.

Список литературы

- Айвазян С. А.* Прикладная статистика: Классификация и снижение размерности. — М.: Финансы и статистика, 1989.
- Иберла К.* Факторный анализ / пер. с нем. В. М. Ивановой; предисл. А. М. Дуброва. — М.: Статистика, 1980.
- Ким Дж.-О., Мьюллер Ч. У., Клекка У. Р.* Факторный, дискриминантный и кластерный анализ. — М.: Финансы и статистика, 1989.
- Кнут Д.* Искусство программирования. Том 2. Получисленные алгоритмы Вильямс, 2000.
- Медынский М. М., Антоний Е. В.* Численные методы нелинейной оптимизации: алгоритмы и программы. — Москва: МАИ, 2003.
- Миркес Е. М.* Нейроинформатика. — Красноярск: ИПЦ КГТУ, 2002.
- Фофанов П. Н.* Учебное пособие по механокардиографии. Военно-медицинская ордена Ленина Краснознаменная академия имени С.М. Кирова, Ленинград, 1977.
- Чазова И. Е.* и др. Диагностика и лечение артериальной гипертензии // Рекомендации Российского медицинского общества по артериальной гипертензии и Всероссийского научного общества кардиологов. — Москва, 2008.