

УДК 629.463

Е. П. ГУРСКИЙ, младший научный сотрудник; Белорусский государственный университет транспорта, г. Гомель

АНАЛИЗ МУЛЬТИКОЛЛИНЕАРНОСТИ ФАКТОРОВ, ОПРЕДЕЛЯЮЩИХ ПОТРЕБНЫЙ РАБОЧИЙ ПАРК ГРУЗОВЫХ ВАГОНОВ

Управление активной частью производственных фондов – подвижным составом, а также формирование инвестиционной программы Белорусской железной дороги, планирование уровня развития постоянных устройств, решение других важных стратегических задач невозможно без научного обоснования потребного парка вагонов. Очевидно, что эта задача чрезвычайно сложна и для ее решения нужен тщательный технико-эксплуатационный анализ и на его основе прогноз потребности в грузовых вагонах. Для решения этой задачи целесообразно использовать селективные алгоритмы и построенные на их основе прогнозные модели оптимальной сложности. Анализируя практические результаты формирования рабочего парка грузовых вагонов, можно утверждать, что они являются следствием воздействия многих факторов, как внешних, так и внутренних, определяющих его структуру и количественное состояние. В связи с этим одним из возможных методов прогнозирования является построение корреляционно-регрессионной модели оптимальной сложности. Важным моментом построения таких моделей при отборе факторов-аргументов является проверка их на мультиколлинеарность. В ходе исследований разработан алгоритм построения прогнозной модели расчета потребного вагонного парка; дана классификация факторов, оказывающих влияние на величину парка вагонов; рассмотрен критерий включения факторов в прогнозную модель – коэффициент корреляции; получены корреляционные матрицы взаимосвязей как факторов и парка вагонов, так и факторов между собой; выполнен мультиколлинеарный анализ.

Сложившаяся ситуация с грузовым подвижным составом в Республике Беларусь остро ставит проблему его оздоровления и обновления. Это связано с тем, что парк вагонов Республики Беларусь стареет. Средний срок службы грузовых вагонов на 01.08.2005 г. составил 21 год, количество вагонов инвентарного парка дороги с истекшим нормативным сроком службы увеличилось до 39,7 % [1]. Устаревший в техническом отношении вагонный парк, износ которого на сегодняшний день уже достиг 74 %, не позволяет перейти на современные перевозочные технологии, устойчиво обеспечивать потребность отрасли в перевозках, увеличить межремонтные сроки, гарантийные плечи пробега, вес поездов, решать многие другие задачи, диктуемые временем, требованиями безопасности движения, снижения эксплуатационных расходов и транспортных издержек. Такое положение дел требует экстренных мер по оздоровлению подвижного состава, повышению эффективности его использования и усилению вагоноремонтной базы. Рациональное, экономически выгодное решение этой задачи невозможно без четкого обоснования исходя из прогнозируемых объемов перевозок, а также ряда внутренних и внешних факторов перспективных потребных парков.

Вопросам перспективного планирования, с целью получения научного прогноза, в последнее время уделяется все большее внимание как у нас в стране, так и за рубежом. Для железнодорожного транспорта в целом характерно возрастание динамики работы всех элементов многоуровневой системы. Важным моментом в этих условиях является существенное возрастание влияния фактора неопределенности на эффективность работы железнодорожного транспорта. В связи с этим ошибки,

допущенные на стадии прогнозирования, повлекут за собой огромные экономические потери и в конечном итоге окажут существенное воздействие на конечные результаты работы транспорта. Эти обстоятельства предъявляют повышенные требования к разработке прогнозных значений основных показателей, характеризующих эффективность работы транспорта. Необходимым условием повышения научного уровня плановой и аналитической работ является использование экономико-математических и математико-статистических методов, позволяющих, наряду с качественным анализом, проводить количественную оценку технико-экономических процессов, происходящих на железнодорожном транспорте, о работе которого ежедневно собирается и обрабатывается огромный объем информации, необходимый для решения оперативных и технико-экономических задач.

Разработка прогнозной модели расчета перспективных значений потребного вагонного парка, поиск рационального их объема – это экономия возможных расходов на закупку подвижного состава, снижение эксплуатационных расходов на его обслуживание и ремонт, выработка обоснованных решений развития собственной базы индустриального ремонта и производства элементов подвижного состава.

Анализ существующих отечественных и зарубежных разработок, а также проведенные исследования возможности применения однофакторных и многофакторных корреляционно-регрессионных моделей позволили установить необходимость применения принципиально новой модели, учитывающей вероятностный характер величин. Анализируя практические результаты формирования рабочего парка грузовых вагонов с 1993 года, можно

утверждать, что они являются следствием воздействия многих факторов, как внешних, так и внутренних, определяющих его структуру и количественное состояние, а именно: грузооборот, пассажирооборот, техническая скорость, участковая скорость, вес поезда, оборот вагона, среднесуточный пробег, статическая нагрузка, динамическая нагрузка, производительность вагона и т. д.

Сложность построения таких моделей заключается в том, что система должна быть способна к гибкому реагированию на изменения, происходящие в ней, т. е. адаптивной. Важнейшим условием точности прогноза является построение адекватной модели выбора наиболее значимых факторов, определяющих закономерности изменения исследуемой величины. Многообразие процессов, под воздействием которых формируются вагонные парки, по своей природе носит весьма сложный характер и определяется многомерной системой некоторых факторов. Поэтому очень ответственным этапом прогнозирования является отбор факторов и определение круга показателей, с помощью которых устанавливается их влияние на исследуемый признак. Многообразие факторов, их взаимодействие, наличие противоположных тенденций (часто неизвестных) влияния их на прогнозируемый показатель приводят к тому, что существующие методы отбора оказываются не всегда работоспособными. Зачастую при исследовании многофакторных моделей искажение действительной зависимости сопоставляемых переменных может быть обусловлено взаимным влиянием факторных переменных, т. е. мультиколлинеарностью. Наличие в модели взаимосвязанных признаков ведет к дублированию влияния одних и тех же факторов на результативный показатель. Вследствие этого коэффициент корреляции между изучаемыми признаками будет завышен.

В практике исследований зачастую встречается вывод о том, что чем больший объем информации по временному ряду используют исследователи, тем более точный результат он получит. Однако это характерно только для стабильных процессов. При нестабильных процессах существует оптимальная длина ретроспективного ряда, при котором не теряется смысл изучаемого явления. Вместе с тем требуется значительно меньший объем работ по хранению, подготовке и обработке исходной информации. Большая длина ряда ретроспекции приводит к тому, что на прогнозные значения сильное влияние оказывает начальная информация, что в конечном итоге приводит к искажению результатов. Необоснованно уменьшенный ряд не позволяет изучить устоявшуюся динамику процесса, что также исказит полученные результаты. Выполненные исследования показали, что рациональная длина ретроспективного ряда – 10 – 12 лет.

В ходе исследований экспертным путем были определены 17 факторов, в той или иной степени

влияющих на формирование вагонного парка, и произведен их статистический анализ за двенадцатилетний период работы с 1993 по 2004 годы. Рассмотрению подверглись следующие факторы: грузооборот – X_1 ; пассажирооборот – X_2 ; участковая скорость – X_3 ; техническая скорость – X_4 ; вес поезда – X_5 ; оборот – X_6 ; среднесуточный пробег – X_7 ; статическая нагрузка – X_8 ; погрузка – X_9 ; вывоз – X_{10} ; коэффициент местной работы – X_{11} ; коэффициент порожнего пробега – X_{12} ; производительность – X_{13} ; динамическая нагрузка – X_{14} ; работа дороги – X_{15} ; простой под грузовой операцией – X_{16} ; простой на одной технической станции – X_{17} . В этом случае алгоритм построения корреляционно-регрессионной модели оптимальной сложности включает в себя следующие этапы:

- выбор основных факторов, определяющих величину парка грузовых вагонов;
- расчет коэффициентов корреляции между исследуемой величиной и факторами, а также факторов между собой;
- расчет стандартизованных коэффициентов множественной регрессии;
- ранжирование факторов по значимости;
- формирование многофакторной модели оптимальной сложности.

Вторым этапом формирования многофакторной корреляционно-регрессионной модели оптимальной сложности является расчет коэффициентов корреляции между исследуемой величиной и факторами, а также у факторов между собой. Расчет коэффициентов парной корреляции производится по формуле [2, 3]

$$r_{kj} = \frac{\sum_{i=1}^n (X_{ki} - \bar{X}_k)(X_{ji} - \bar{X}_j)}{(n-1)\sigma_k\sigma_j} \quad (1)$$

$$j = \overline{1, M}; \quad k = \overline{1, M}.$$

Наличие массива коэффициентов парной корреляции позволяет рассчитать совокупный коэффициент корреляции

$$R = \sqrt{\left| \frac{\Delta_1}{\Delta} \right|}, \quad (2)$$

$$\Delta_1 = \begin{vmatrix} r_{12} & r_{13} & \dots & r_{1M} & 0 \\ 1 & r_{23} & \dots & r_{2M} & r_{21} \\ r_{32} & 1 & \dots & r_{3M} & r_{31} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ r_{M2} & r_{M3} & \dots & 1 & r_{M1} \end{vmatrix}, \quad (3)$$

$$\Delta = \begin{vmatrix} 1 & r_{23} & r_{24} & \dots & r_{2M} \\ r_{32} & 1 & r_{34} & \dots & r_{3M} \\ r_{42} & r_{43} & 1 & \dots & r_{4M} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ r_{M2} & r_{M3} & r_{M4} & \dots & 1 \end{vmatrix}$$

Совокупный коэффициент корреляции, равный 1,0, свидетельствует о том, что выбранные факторы в полной мере определяют величину рабочего парка грузовых вагонов.

В различных источниках теснота связи исследуемых факторов по рассчитанному коэффициенту корреляции оценивается по-разному. Наиболее приемлемая оценка дана в [2], где указывается, что при определенных условиях между исследуемыми величинами существует следующая связь:

$0 \leq |r| \leq 0,20$ – практически отсутствует;

$0,20 \leq |r| \leq 0,50$ – слабая;

$0,50 \leq |r| \leq 0,75$ – средняя;

$0,75 \leq |r| \leq 0,95$ – сильная;

$0,95 \leq |r| \leq 1,00$ – практически функциональная.

Методы парной и множественной корреляции в настоящее время занимают большое место в теории прогнозирования. Они обладают рядом существенных достоинств: сравнительная простота реализации на ЭВМ, легкость получения прогнозных зависимостей, незначительные затраты труда на подготовку исходной информации, высокая степень объективности результатов и др.

Однако наряду с известными преимуществами эти модели обладают и существенными недостатками. Одним из главных является довольно часто встречающаяся коррелируемость факторов друг с другом. Чем сильнее эта связь, тем в меньшей сте-

пени проявляется влияние их на прогнозируемую величину. Все это в конечном итоге оказывает решающее влияние на устойчивость коэффициентов линий регрессии. В этом случае добавление или уменьшение объема многофакторной модели хотя бы на один фактор может сильно изменить прогнозную модель. При этом зачастую изменяются не только величины коэффициентов регрессии, но и направление их влияния. Зачастую мультиколлинеарность проявляется не по всей совокупности факторов, а только для отдельных групп. Это в еще большей степени усложняет решение задачи.

Как было отмечено ранее, за критерий целесообразности включения фактора в модель был принят коэффициент корреляции. Если при включении фактора коэффициент растет, то причинная связь признается существенной. Одним из основных требований включения фактора в модель является отсутствие взаимной корреляции у факторов между собой. На этом этапе из модели удаляются те факторы, которые имели достаточно большой ($r \geq 0,75$) коэффициент корреляции с оставшимися в модели факторами. В модели остается фактор, который носит более общий характер и имеет больший коэффициент корреляции между исследуемой величиной и факторами.

Результат расчета межфакторных коэффициентов корреляции приведен в таблице 1.

Таблица 1 – Корреляционная матрица

	$N_{\text{раб}}$ (Y)	$\sum Pl$ (X_1)	$\sum al$ (X_2)	$V_{\text{уч}}$ (X_3)	$V_{\text{вс}}$ (X_4)	$P_{\text{поезд}}$ (X_5)	O (X_6)	$t_{\text{сп.свт.}}$ (X_7)	$P_{\text{ст.}}$ (X_8)	$P_{\text{погр.}}$ (X_9)	$\sum Pl_{\text{ваг}}$ (X_{10})	k_m (X_{11})	$k_{\text{погр.}}$ (X_{12})	$W_{\text{сп.свт.}}$ (X_{13})	$P_{\text{л}}$ (X_{14})	U (X_{15})	$t_{\text{сп.опер.}}$ (X_{16})	$t_{\text{вс.ст.}}$ (X_{17})
$N_{\text{раб}}$ (Y)	1	0,89	0,031	0,382	-0,264	0,843	0,819	-0,702	0,696	0,589	0,915	-0,881	0,645	-0,675	0,28	0,812	0,89	0,882
$\sum Pl$ (X_1)	0,89	1	0,07	0,293	-0,453	0,862	0,511	-0,309	0,704	0,79	0,907	-0,769	0,764	-0,268	0,284	0,95	0,661	0,801
$\sum al$ (X_2)	0,031	0,07	1	0,459	0,36	0,292	0,086	0,015	0,292	0,321	0,19	-0,202	0,045	0,11	0,614	-0,021	0,249	-0,251
$V_{\text{уч}}$ (X_3)	0,382	0,293	0,459	1	0,458	0,61	0,563	-0,436	0,751	0,502	0,511	-0,608	0,581	-0,326	0,76	0,102	0,599	0,111
$V_{\text{вс}}$ (X_4)	-0,264	-0,453	0,36	0,458	1	-0,301	0,011	-0,138	-0,206	-0,078	-0,239	0,077	-0,349	-0,125	0,144	-0,423	-0,008	-0,46
$P_{\text{поезд}}$ (X_5)	0,843	0,862	0,292	0,61	-0,301	1	0,716	-0,487	0,95	0,749	0,927	-0,847	0,828	-0,393	0,647	0,669	0,833	0,644
O (X_6)	0,819	0,511	0,086	0,563	0,011	0,716	1	-0,954	0,694	0,261	0,704	-0,829	0,468	-0,913	0,436	0,333	0,96	0,651
$t_{\text{сп.свт.}}$ (X_7)	-0,702	-0,309	0,015	-0,436	-0,138	-0,487	-0,954	1	-0,462	-0,05	-0,515	0,668	-0,218	0,988	-0,281	-0,183	-0,864	-0,559
$P_{\text{ст.}}$ (X_8)	0,696	0,704	0,292	0,751	-0,206	0,95	0,694	-0,462	1	0,661	0,818	-0,793	0,851	-0,348	0,76	0,46	0,768	0,505
$P_{\text{погр.}}$ (X_9)	0,589	0,79	0,321	0,502	-0,078	0,749	0,261	-0,05	0,661	1	0,703	-0,588	0,615	0,044	0,569	0,726	0,479	0,33
$\sum Pl_{\text{ваг}}$ (X_{10})	0,915	0,907	0,19	0,511	-0,239	0,927	0,704	-0,515	0,818	0,703	1	-0,823	0,763	-0,459	0,421	0,792	0,813	0,782
k_m (X_{11})	-0,881	-0,769	-0,202	-0,608	0,077	-0,847	-0,829	0,668	-0,793	-0,588	-0,823	1	-0,731	0,616	-0,431	-0,616	-0,9	-0,717
$k_{\text{погр.}}$ (X_{12})	0,645	0,764	0,045	0,581	-0,349	0,828	0,468	-0,218	0,851	0,615	0,763	-0,731	1	-0,155	0,385	0,611	0,54	0,6
$W_{\text{сп.свт.}}$ (X_{13})	-0,675	-0,268	0,11	-0,326	-0,125	-0,393	-0,913	0,988	-0,348	0,044	-0,459	0,616	-0,155	1	-0,128	-0,177	-0,805	-0,579
$P_{\text{л}}$ (X_{14})	0,28	0,284	0,614	0,76	0,144	0,647	0,436	-0,281	0,76	0,569	0,421	-0,431	0,385	-0,128	1	0,048	0,529	-0,041
U (X_{15})	0,812	0,95	-0,021	0,102	-0,423	0,669	0,333	-0,183	0,46	0,726	0,792	-0,616	0,611	-0,177	0,048	1	0,491	0,777
$t_{\text{сп.опер.}}$ (X_{16})	0,89	0,661	0,249	0,599	-0,008	0,833	0,96	-0,864	0,768	0,479	0,813	-0,9	0,54	-0,805	0,529	0,491	1	0,668
$t_{\text{вс.ст.}}$ (X_{17})	0,882	0,801	-0,251	0,111	-0,46	0,644	0,651	-0,559	0,505	0,33	0,782	-0,717	0,6	-0,579	-0,041	0,777	0,668	1

В литературе зачастую встречаются рекомендации о том, чтобы не принимать к дальнейшим расчетам факторы, имеющие низкий коэффициент корреляции с исследуемой величиной, в данном случае – $N_{\text{раб}}$. Однако это следует отнести только к однофакторным моделям. Мы же имеем дело с многофакторными моделями. В этом случае мы не рекомендуем снимать с анализа ни одного фактора. Например, при формировании модели рабочего парка платформ $N_{\text{раб.платф}} = f(X_4)$ коэффициент корреляции $r_{N_{\text{раб.платф}}, X_4 (\text{оборот})} = 0,66$. А далее для моделей

$$N_{\text{раб.платф}} = f(X_4, X_6) -$$

$$- r_{N_{\text{раб.платф}}, (X_4, X_6 (\text{работа}))} = 0,92;$$

$$N_{\text{раб.платф}} = f(X_4, X_6, X_5) -$$

$$- r_{N_{\text{раб.платф}}, (X_4, X_6, X_5 (\text{погр.строит.грузов}))} = 0,95.$$

На этом шаге формирование модели можно прекратить, так как $r_{N_{\text{раб.платф}}, (X_4, X_6, X_5)} = 0,95$. Если следовать литературным рекомендациям, то при $r_{N_{\text{раб.платф}}, X_4} = 0,66$; $r_{N_{\text{раб.платф}}, X_6} = -0,23$; $r_{N_{\text{раб.платф}}, X_5} = 0,33$ эти факторы из дальнейшего анализа следовало бы исключить.

Целью анализа взаимных коэффициентов корреляции является установление и после глубокого анализа исключение одного из факторов, обладающего высоким уровнем взаимного коэффициента корреляции. Как было сказано выше, из двух

факторов с таким уровнем $r_{(x_i, x_j)}$ для дальнейшего рассмотрения остается фактор более общего характера и имеющий больший $r_{N_{\text{раб}}, x_i}$. По полученным результатам расчета следует рассмотреть $r_{x_1, x_5} = 0,86$, $r_{x_1, x_9} = 0,79$, $r_{x_1, x_{10}} = 0,90$, $r_{x_1, x_{11}} = -0,77$, $r_{x_1, x_{12}} = 0,76$, $r_{x_1, x_{15}} = 0,95$, $r_{x_1, x_{17}} = 0,80$, $r_{x_3, x_8} = 0,75$, $r_{x_3, x_{14}} = 0,76$, $r_{x_6, x_7} = -0,95$, $r_{x_6, x_{13}} = -0,91$, $r_{x_6, x_{16}} = 0,96$ и т. д.

Графическая интерпретация анализа мультиколлинеарности факторов представлена на рисунке 1.

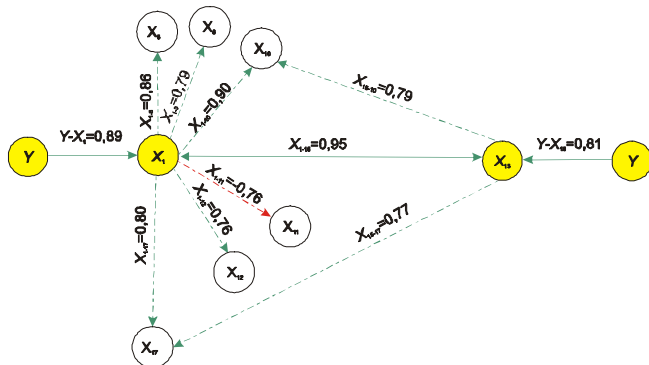


Рисунок 1 – Анализ межфакторных коэффициентов корреляции

Анализ расчета коэффициентов корреляции позволяет отметить наличие довольно тесных корреляционных связей: грузооборот – работа, рабочий парк – грузооборот, рабочий парк – работа, с коэффициентами корреляции соответственно $r_{x_1, x_{15}} = 0,95$, $r_{N_{\text{раб}}, x_1} = 0,89$, $r_{N_{\text{раб}}, x_{15}} = 0,81$. Грузооборот, являясь важнейшим системным фактором в модели, носит более общий характер. С ним также тесно связаны такие факторы, как вес поезда, погрузка, вывоз, коэффициент местной работы, коэффициент порожнего пробега, простой на одной технической станции. Коэффициент корреляции $r_{N_{\text{раб}}, x_1} > r_{N_{\text{раб}}, x_{15}}$, поэтому исходя из двух условий включения факторов в модели оставляем грузооборот. Аналогично производится анализ остальных межфакторных связей.

В результате анализа в многофакторной модели расчета общего рабочего парка грузовых вагонов

Получено 17.03.2006

E. P.Gursky. Analysis multikollinearности between factor, defining required a park freiht-car.

Control active part production assets – a rolling stocks, as well as shaping the investment program of the Belorussian railway, planning level developments constant device, decision other important strategic tasks impossible without scientific motivation required coach parka. Obviously that this task exceedingly complex and for its decisions needs the careful technician-working analysis and use on their base forecast need for freiht-cars. For decision of this task reasonable to use the selective algorithms and built on their base прогнозные to models to optimum difficulty. Analysing practical results of the shaping worker parka of the freiht-cars, possible confirm that they are due to influences many factor, as external, so and internal, defining structure and quantitative condition. A building In this connection one of the possible methods of the forecasting correlation models to optimum difficulty. The Important moment of the building of such models, at selection factor-argument, is a check them on multicorrelation. In the course of studies is designed algorithm of the building forecastion to models of the calculation required rolling stock; it is given categorization factor, influencing on value parka coach; the considered criterion of the cut-in factor in forecastion model - a factor to correlations; they are received correlation matrixes of the inter-coupling as factor and parka coach, so and factor between itself; was executed multicorrelation analysis.

были оставлены следующие факторы: X_1 – грузооборот; X_2 – пассажирооборот; X_3 – участковая скорость; X_4 – техническая скорость; X_5 – оборот.

Наличие коэффициентов корреляции между рабочим парком и факторами позволяет выполнить их ранжирование по степени значимости. Фактору с наибольшим коэффициентом корреляции присваивается один балл и т. д. При ранжировании по коэффициенту корреляции факторы расположились следующим образом

$$\begin{aligned} X_1 (r_{N_{\text{раб}}, x_1} = 0,89) &\rightarrow X_5 (r_{N_{\text{раб}}, x_5} = 0,81) \rightarrow \\ &\rightarrow X_3 (r_{N_{\text{раб}}, x_3} = 0,38) \rightarrow X_4 (r_{N_{\text{раб}}, x_4} = -0,26) \rightarrow \\ &\rightarrow X_2 (r_{N_{\text{раб}}, x_2} = 0,03). \end{aligned}$$

Таким образом, важное значение в совершенствовании системы стратегического управления рабочим парком вагонов имеет корреляционный анализ влияния различных факторов на рабочий парк вагонов. Выполненные исследования и полученные результаты позволили установить регрессионные зависимости и корреляционные взаимосвязи рассматриваемых показателей с рабочим парком грузовых вагонов. Полученные результаты позволяют на начальном этапе исследования предварительно оценить степень влияния факторов на исследуемую величину.

Список литературы

- 1 **Сенько, В. И.** Вопросы качественного и количественного состояния инвентарного парка грузовых вагонов в Республике Беларусь / В. И. Сенько, Е. П. Гурский // Сучасні проблеми та ефективні шляхи ремонту і відновлення залізничного рухомого складу: сб. тез. науч.-практ. конф. Киев, 18–19 мая 2006 г. – Киевгипротранс, 2006. – С. 13–14.
- 2 **Правдин, Н. В.** Прогнозирование грузовых потоков / Н. В. Правдин, М. Л. Дыканюк, В. Я. Негрей. – М.: Транспорт, 1987. – 247 с.
- 3 **Сенько, В. И.** Совершенствование организации технического обслуживания и текущего ремонта грузовых вагонов / В. И. Сенько. – Гомель: БелГУТ, 2002. – 178 с.
- 4 **Кобринский, Н. Е.** Точность экономико-математических моделей / Н. Е. Кобринский, В. И. Кузьмин. – М.: Финансы и статистика, 1981. – 255 с.
- 5 **Кильдешев, В. А.** Анализ временных рядов и прогнозирование / В. А. Кильдешев, Л. М. Френкель. – М.: Статистика, 1973. – 103 с.