ПРИМЕНЕНИЕ ВЕРОЯТНОСТНЫХ МЕТОДОВ ПРИ ОПРЕДЕЛЕНИИ ПРОДОЛЖИТЕЛЬНОСТИ РОСПУСКА СОСТАВОВ С ГОРКИ

Н. Е. Гончаров
Киев

При разработке системы автоматизации планирования поездообразования на одной из сортировочных станций Юго-Западной железной дороги необходимо было установить аналитические выражения зависимостей продолжительности роспуска составов с горки от различных факторов.

Поскольку разработка и внедрение системы автоматизации поездообразования на этой станции намечалось в условиях работы станции без существенного изменения действующей технологии — до внедрения устройств автоматизации роспуска составов (АРС), без автомашиниста, автосцепщика вагонов, автоматического определения высот осей автосцепок и др.,— на первом этапе предлагается исследовать зависимость продолжительности роспуска составов от различных факторов, имеющих место в реальных условиях.

С этой целью было прохронометрировано расформирование с горки около 400 составов грузовых поездов. При этом фиксировались такие элементы, как количество вагонов в составе (физических и учетных), количество отцепов, длина составов и отцепов, начало и окончание надвига и роспуска, их продолжительность, количество остановок за этот период, средняя скорость роспуска.

Предварительный анализ показал, что для различных составов средняя скорость и время роспуска переменны, а величина их зависит от многих факторов и прежде всего от длины составов и числа отцепов. С увеличением длины составов время роспуска, как правило, увеличивается, но очень часто, даже при одинаковой длине составов, затраты времени на роспуск оказываются различными. Причем, время роспуска увеличивается с ростом числа отцепов в составе, так как появляется необходимость снижения скорости роспуска.

Анализ данных хронометражных наблюдений, показал, что наиболее четко проявляется определенная зависимость затраты времени, приходящейся на один отцеп, от числа вагонов в каждом отцепе. Это подтверждают точки на рис. 1 — затраты времени на отцеп ($t_{от}$) при соответствующем числе вагонов в отцепе ($m_{от}$), полученные в результате хронометражных наблюдений.

Анализируя размещения точек, видим, что они группируются крайне неравномерно. Наибольшее их число приходится от одного до семи

вагонов в отцепе, в указанных пределах частота их размещения различна. Частота отцепов по числу вагонов (учетных) в каждом отцепе приведена в следующей табл. 1.

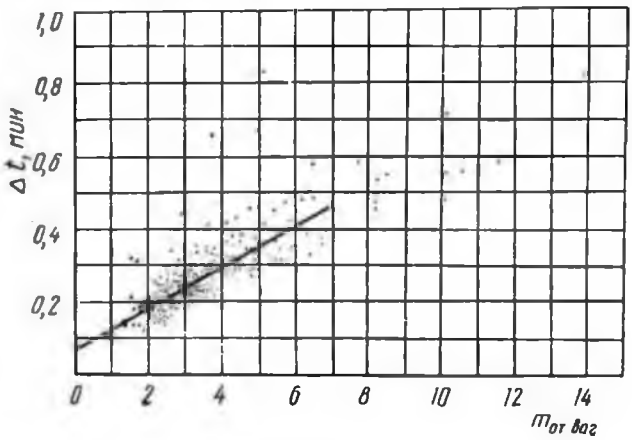

Рис. 1.

Таблица 1

Количество вагонов в отцепе ($m_{от}$)	1—1,5	1,5—2	2—2,5	2,5—3	3—3,5	3,5—4	4—4,5	4,5—5	5—5,5	5,5—6	6—6,5	6,5—7
Количество отцепов (частота) P_i	6	21	49	69	61	44	32	21	12	7	4	2

По данным интервального вариационного ряда, представленного в таблице, строится гистограмма (рис. 2).

Далее необходимо построить полигон, для чего данный интервальный ряд преобразуется в дискретный. При этом каждый интервал заменяется значением признака, приходящимся на его средину, и составляется таблица, по данным которой строится полигон (рис. 3).

Таблица 2

$m_i^{от}$	1,25	1,75	2,25	2,75	3,25	3,75	4,25	4,75	5,25	5,75	6,25	6,75
P_i	6	21	49	69	61	44	32	21	12	7	4	2

Предположим, что полученное эмпирическое распределение является нормальным, т. е. может быть описано дифференциальной функцией

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \cdot e^{-\frac{(x-a)^2}{2\sigma^2}},$$

где $x = m_i^{от}$;

$a = \overline{m}_{от}$ — математическое ожидание;

σ — среднее квадратическое отклонение нормально распределенной случайной величины.

Если обозначить

$$u_i = \frac{m_i^{\text{от}} - \bar{m}_{\text{от}}}{\sigma};$$

$$\varphi(u_i) = \frac{1}{2\pi} \cdot e^{-\frac{u_i^2}{2}},$$

то выравнивающие частоты, ординаты теоретической кривой, определяются по формуле

$$y_i = \frac{P \cdot h}{\sigma} \varphi(u_i),$$

где P — сумма наблюденных частот;

h — разность между двумя соседними вариантами.

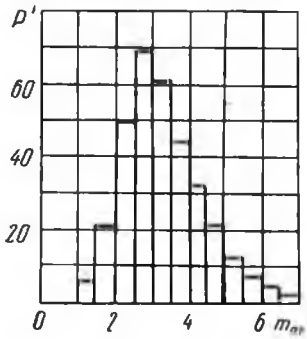

Рис. 2.

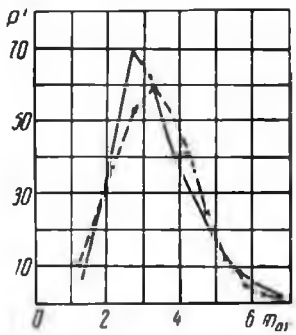

Рис. 3.

Величины P, h введены в исходную формулу, так как определяются не относительные, а абсолютные значения частот.

Необходимо построить нормальную кривую, исходя из приведенных опытных данных, и проверить статическую гипотезу о нормальном распределении генеральной совокупности.

Построение нормальной кривой производится в следующей последовательности.

1) Составляется расчетная вспомогательная табл. 3.

2) По итоговым данным граф 3 и 4 этой таблицы определяется математическое ожидание

$$m_{\text{от}} = \frac{\Sigma m_i^{\text{от}} \cdot P_i}{\Sigma P_i} = \frac{1085}{328} = 3{,}3 \text{ вагона.}$$

3) Дисперсия, характеризующая рассеивание случайных величин относительно математического ожидания, рассчитывается по формуле

$$\sigma^2 = \frac{\Sigma (m_i^{\text{от}} - \bar{m}_{\text{от}})^2 \cdot P_i}{\Sigma P_i} = \frac{376}{328} = 1{,}15.$$

Значения принимаются по итоговым данным граф 3 и 7 табл. 3.

4) Значения u_i определяются в графе 11 вспомогательной табл. 3, значения $\varphi(u_i)$ — в приложении *.

---

* В. Е. Гмурман. Введение в теорию вероятностей и математическую статистику. Изд-во «Высшая школа», 1966.

Таблица 3

Расчетная вспомогательная

Интервалы (вагоны)	Центры интервалов $m_{от}$	Эмпиричность частоты	$m_{от} P_i$	$m_{от} - \bar{m}_{от}$	$(m_{от} - \bar{m}_{от})^2$	$(m_{от} - \bar{m}_{от})^2 P_i$	$(m_{от} - \bar{m}_{от})^3 \cdot P_i$	$(m_{от} - \bar{m}_{от})^4 \cdot P_i$	Накопление частоты S	$u_i = \frac{m_{от} - \bar{m}_{от}}{\sigma}$	$\varphi(u_i)$	P'_i	$P_i - P'_i$	$(P_i - P'_i)^2$	$\frac{(P_i - P'_i)^2}{P'_i}$
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
1,0—1,5	1,25	6	7,5	—2,05	4,25	25,5	—52,2	108,2	6	—1,91	0,0644	10	—4	16	1,60
1,5—2,0	1,75	21	36,8	—1,55	2,40	50,4	—78,0	121,0	27	—1,45	0,1349	22	—1	1	0,04
2,0—2,5	2,25	49	110,3	—1,05	1,10	53,9	—56,5	59,2	76	—0,98	0,2468	38	+11	121	3,20
2,5—3,0	2,75	69	189,5	—0,55	0,30	20,5	—11,3	6,2	145	—0,51	0,3503	54	+15	225	4,22
3,0—3,5	3,25	61	198,3	—0,05	—	—	—	—	206	—0,05	0,3984	62	—1	1	0,02
3,5—4,0	3,75	44	165,0	0,45	0,20	8,8	4,0	1,8		0,42	0,3653	56	—12	144	2,58
4,0—4,5	4,25	32	136,0	0,95	0,90	28,8	27,4	26,0		0,89	0,2685	41	—9	81	1,97
4,5—5,0	4,75	21	99,8	1,45	2,10	44,0	63,7	92,3		1,35	0,1604	25	—4	16	0,64
5,0—5,5	5,25	12	63,0	1,95	3,80	45,5	89,0	173,4		1,82	0,0761	13	—1	1	0,08
5,5—6,0	5,75	7	40,3	2,45	6,00	42,0	102,7	252,2		2,39	0,0229	4	—3	9	2,25
6,0—6,5	6,25	4	25,0	2,95	8,70	349,8	102,6	302,5		2,76	0,0088	2	—2	4	2,00
6,5—7,0	6,75	2	13,5	3,45	10,95	21,8	75,6	240,5		3,23	0,0022	1	—1	1	1,00
		328	1085,0			376,0	267,0	1383,0							19,6

5) В графе 13 рассчитываются значения ординат теоретической нормальной кривой

$$P' = \frac{P}{\sigma}\varphi(u_i) = \frac{328 \cdot 0{,}5}{1{,}07} \cdot \varphi(u_i) = 153.$$

6) Точки $(m_i^{от}, P_i')$ наносятся на полигон (рис. 3) и соединяются пунктирной линией. Близость выравнивающих частот к наблюдаемым подтверждает правильность допущения о том, что исследуемый признак распределен нормально.

Однако во избежание противоречивых заключений, следует проверить гипотезу о нормальном распределении с помощью критерия согласия, который определяет, является ли расхождение эмпирического и теоретического распределений случайным, т. е. определить, согласуются ли данные наблюдений с выдвинутой статистической гипотезой.

В данной работе проверка производится по критерию Пирсона χ^2:

$$\chi^2 = \frac{(P_i - P_i')^2}{P_i'},$$

где P_i — эмпирические частоты;

P_i' — выравнивающие частоты.

При большем согласовании эмпирического и теоретического распределений меньше различаются эмпирические и выравнивающие частоты и меньше значение χ^2.

Функция распределения зависит также от параметра K, называемого числом степеней свободы. При проверке гипотезы о нормальном распределении $K = S - 3$, где S — число интервалов, на которые разбиты наблюдения. Имеются готовые таблицы [1] (приложение 5), в которых указана вероятность того, что в результате влияния случайных факторов величина χ^2 примет значение, не меньшее, чем вычисленное по данным выборки число χ^2.

Чтобы найти сходимость распределений, необходимо взять определенный уровень значимости, определяемый достаточно малой вероятностью, при которой (в данной определенной задаче) событие можно считать практически невозможным. На практике принимают уровни значимости в пределах 0,01—0,05 (одно-пятипроцентные).

В данной задаче уровень значимости 0,01. Если вероятность, найденная по таблицам, окажется меньше 0,01, это означает, что в результате влияния случайных причин наступило событие, которое в силу принципа практической невозможности маловероятных событий практически невозможно. Если вероятность, найденная по таблицам, больше 0,01, гипотеза о нормальном распределении генеральной совокупности согласуется с данными наблюдений, и поэтому может быть принята.

Расчет выполняется в следующей последовательности.

1) вычисляются выравнивающие частоты (графа 13, табл. 3);

2) вычисляется «квадрат Пирсона» (графы 14, 15, 16, табл. 3).

$$\chi^2 = 19{,}6;$$

3) определяется число степеней свободы

$$K = 12 - 3 = 9;$$

4) выбирается уровень значимости 0,01;

5) по таблице работы [1] (приложение 5) определяется значение вероятности $P\ (\chi^2 \geqslant \chi_q^2)$, равное после интерполяции и экстраполяции 0,017.

Поскольку это значение оказалось больше уровня значимости 0,01, гипотеза о нормальном распределении рассмотренной генеральной совокупности принимается.

Для более точной оценки отклонения эмпирического распределения от нормального используются различные характеристики, к числу которых относятся: асимметрия r_3 и аксцесс E.

Асимметрия (скошенность) вариационного ряда определяется по формуле

$$r_3 = \frac{\mu_3}{\sigma^3}; \quad \mu_3 = \frac{(\Sigma m_i^{от} - \bar{m}_{от})^3 \cdot P_i}{\Sigma P_i};$$

$$\sigma^3 = 1{,}23: \quad r_3 = \frac{267}{1{,}23 \cdot 328} = 0{,}67.$$

Небольшое положительное число свидетельствует о небольшой правосторонней асимметрии этого распределения.

Эксцесс (крутость) вариационного ряда определяется по формуле

$$E = r_4 - 3,$$

где $r_4 = \frac{\mu_4}{\sigma^4}$ — мера крутости;

$\mu_4 = \frac{\Sigma(m^{от} - \bar{m}_{от})^4 \cdot P_i}{\Sigma P_i}$ — центральный момент 4-го порядка;

$$\mu_4 = \frac{1383}{319} = 4{,}35; \quad r_4 = \frac{4{,}35}{1{,}32} = 3{,}3;$$

$$E = 3{,}3 - 3 = 0{,}3.$$

Полученная положительная величина эксцесса свидетельствует о высоковершинности распределения, которое немного отличается от нормального.

Далее рассчитываются значения медианы M_e, моды M_o и вариационного размаха R.

Поскольку данный рассматриваемый ряд является интервальным, то медиана (срединное значение или вероятное значение) определяется по формуле

$$M_e = m_0^{от} + \frac{\frac{\Sigma P_i}{2} - S_{M_e-1}}{P_{M_e}} \cdot k,$$

где $m_0^{от}$ — начало медианного интервала;

ΣP_i — сумма всех частот;

M_{e-1} — сумма частот до медианного интервала;

P_{M_e} — частота медианного интервала;

k — ширина медианного интервала.

Для нахождения медианного интервала складываются последовательно частоты и записываются нарастающим итогом.

Первой накопленной частоте S_{Me}, большей $0{,}5\Sigma P_i$, соответствует медианный интервал.

Из табл. 1 видно, что $0{,}5\,\Sigma P_i = 164$, тогда медианным будет интервал 3—3,5, при котором справедливы следующие исходные данные:

$$m_0^{от} = 3{,}0; \quad k = 0{,}5; \quad S_{M_e-1} = 145; \quad P_{M_e} = 61; \quad \Sigma P_i = 328;$$

БИБЛИОТЕКА
Инв. 520062

$$M_e = 3{,}0 + \frac{164 - 145}{61}\,0{,}5 = 3{,}16 \text{ вагона.}$$

По данным рассматриваемого упорядоченного интервального вариационного ряда мода (наивероятнейшее значение) определяется по формуле

$$M_0 = m_0^{от} + \frac{P_{M_0} - P_{M_0-1}}{(P_{M_0} - P_{M_0-1}) + (P_{M_0} - P_{M_0+1})} \cdot k,$$

где $m_0^{от}$ — начало модального интервала;
P_{M_0-1} — частота интервала, предшествующего модальному;
P_{M_0} — частота модального интервала;
P_{M_0+1} — частота интервала, следующего за модальным;
k — ширина модального интервала.

При нахождении моды по данным дискретного ряда за моду принимается значение признака, которому соответствует наибольшая частота. В рассматриваемом случае модальным является интервал 2,5—3, следовательно, $m_0^{от} = 2{,}5$;

$$P_{M_0} = 69; \quad P_{M_0-1} = 49; \quad P_{M_0+1} = 61; \quad k = 0{,}5.$$

$$M_0 = 2{,}5 + \frac{(69 - 49)}{(69 - 49) + (69 - 61)}\,0{,}5 = 2{,}86 \text{ вагона.}$$

Вариационный размах R находим как разность между наибольшим и наименьшим значениями признака в данном вариационном ряду

$$R = m_{от}^{max} - m_{от}^{min} = 6{,}75 - 1{,}25 = 5{,}5 \text{ вагона.}$$

Наиболее вероятное значение числа вагонов в отцепе, соответствующее величине моды,— около трех. При дисперсии 1,15 вагона и среднем квадратическом отклонении 1,07 вагона можно считать, что отцепы величиной 2,5—3,5 вагона встречаются чаще других.

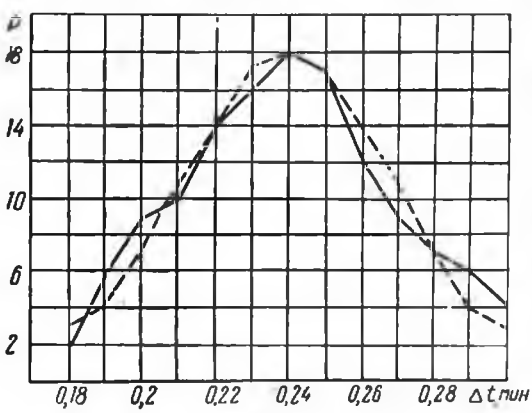

Рис. 4.

Рассмотрим затраты времени, приходящиеся на роспуск одного отцепа, среднее количество вагонов в котором может колебаться в пределах 2,5—3,5. Частота распределения таких отцепов P_i в зависимости от удельной затраты времени на роспуск Δt_i приведена в следующей таблице.

На рис. 4 представлен полигон данного дискретного

Таблица 4

Δt_i	0,18	0,19	0,20	0,21	0,22	0,23	0,24	0,25	0,26	0,27	0,28	0,29	0,30
P_i	2	6	9	10	14	16	18	17	12	9	7	6	4

ряда. Из предварительного анализа следует предположение, что рассматриваемое эмпирическое распределение является нормальным. Что-

бы убедиться в этом, строится нормальная кривая в последовательности, рассмотренной выше.

1. Составляется расчетная вспомогательная табл. 5.

2. По итоговым данным граф 2 и 3 определяется математическое ожидание

$$\Delta\bar{t} = \frac{\Sigma \Delta t_i P_i}{\Sigma P_i} = \frac{31{,}18}{130} = 0{,}24 \text{ мин.}$$

3. По итоговым данным граф 2 и 6 рассматривается дисперсия

$$\sigma^2 = \frac{\Sigma (\Delta t_i - \Delta\bar{t}) P_i^2}{\Sigma P_i} = \frac{0{,}1080}{130} = 0{,}00084.$$

4. Рассчитывается среднее квадратическое отклонение

$$\sigma = \sqrt{0{,}00084} = 0{,}03.$$

5. В графах 7, 8 и 9 определяются значения u_i, $\varphi(u_i)$, а также

$$P' = \frac{Ph}{\sigma} \cdot \varphi(u_i) = \frac{130 \cdot 0{,}01}{0{,}03} \varphi(u_i) = 153\varphi(u_i).$$

6. Полученные значения теоретической нормальной кривой наносятся на полигон (рис. 4) и соединяются пунктирной линией.

Близость выравнивающих частот к наблюдаемым подтверждает правильность допущения о том, что исследуемый признак распределен нормально, о чем свидетельствует проверка по критерию Пирсона.

В графах 10, 11 и 12 вычисляются значения $\chi^2 = 3{,}97$. При числе степеней свободы $K = 13 - 3 = 10$ и уровне значимости 0,01 значение вероятности более 0,85, что говорит о нормальном распределении рассматриваемой генеральной совокупности.

Мода (наивероятнейшее значение) для данного вариационного ряда совпадает со средней арифметической

$$M_0 = 0{,}24 \text{ } мин.$$

В результате выполненных исследований и расчетов определяем наиболее вероятное число вагонов в отцепе — примерно 3 вагона, при этом наиболее вероятная затрата времени при роспуске составов, приходящаяся на один такой отцеп, равна 0,24 *мин*.

При числе вагонов в отцепах в пределах 1—7 время на роспуск каждого отцепа характеризуется точками на рис. 1, причем, по расположению точек можно утверждать наличие прямолинейной зависимости между числом вагонов в отцепах и затратой времени на один отцеп при роспуске составов на сортировочной горке.

Поэтому на рис. 1 наносится точка с координатами $m_{от} = 3$ вагона, $\Delta t = 0{,}24$ *мин* как наиболее вероятная, и через эту точку проводится прямая таким образом, чтобы остальные точки оказались расположенными симметрично относительно этой прямой.

Тогда такая прямая может быть представлена исходным уравнением вида

$$y = ax + c,$$

где $y = \Delta t$ мин;

$x = m_{от}$ вагонов.

Значения коэффициентов уравнения определим из графика (рис. 1). Рассматриваемая прямая отсекает на оси ординат отрезок 0,07 *мин*, что соответствует величине свободного члена исходного уравнения. Для определения значения коэффициента a на оси абсцисс

Таблица 5

Δt_i	P_i	$\Delta t_i P_i$	$\Delta t_i - \Delta\bar{t}$	$(\Delta t_i - \Delta\bar{t})^2$	$(\Delta t_i - \Delta\bar{t})^2 \cdot P_i$	$u_i = \frac{\Delta t_i - \Delta\bar{t}}{\sigma}$	$\varphi(u_i)$	P_i'	$P_i - P_i'$	$(P_i - P_i')^2$	$\frac{(P_i - P_i')^2}{P_i'}$
1	2	3	4	5	6	7	8	9	10	11	12
0,18	2	0,36	—0,06	0,0036	0,0072	—2,00	0,0540	3	—1	1	0,30
0,19	6	1,14	—0,05	0,0025	0,0150	—1,67	0,0989	4	2	4	1,00
0,20	9	1,80	—0,04	0,0016	0,0144	—1,33	0,1647	7	2	4	0,57
0,21	10	2,10	—0,03	0,0009	0,0090	—1,00	0,2420	11	—1	1	0,09
0,22	14	3,08	—0,02	0,0004	0,0056	—0,67	0,3187	14	0	0	0
0,23	16	3,86	—0,01	0,0001	0,0016	—0,33	0,3778	17	—1	1	0,06
0,24	18	4,32	0	0	0	0	0,3989	18	0	0	0
0,25	17	4,25	0,01	0,0001	0,0017	0,33	0,3778	17	0	0	0
0,26	12	3,12	0,02	0,0004	0,0048	0,67	0,3187	14	—2	4	0,29
0,27	9	2,43	0,03	0,0009	0,0081	1,00	0,2420	11	—2	4	0,36
0,28	7	1,96	0,04	0,0016	0,0112	1,33	0,1647	7	0	0	0
0,29	6	1,74	0,05	0,0025	0,0150	1,67	0,0989	4	2	4	1,00
0,30	4	1,20	0,06	0,0036	0,0144	2,00	0,0540	3	1	1	0,30
	130	31,18			0,1080						3,97

можно взять произвольный отрезок, например, $m_{от} = 5$, тогда $\Delta t = 0{,}35$ *мин*. После подстановки этих значений в исходное уравнение

$$\Delta t = 0{,}056 m_{от} + 0{,}07. \quad (1)$$

Общее время на роспуск с горки состава, состоящего из отцепов в среднем до 7 вагонов:

$$t_p = \Delta\, \mathrm{tg}, \quad (2)$$

где g — число групп вагонов в составе.

По данным хронометражных наблюдений имеется около 10% составов с числом вагонов в отцепах более семи. Как показали исследования, в этих случаях существует другая зависимость. На основании анализа данных хронометражных наблюдений на станции *Н* установ-

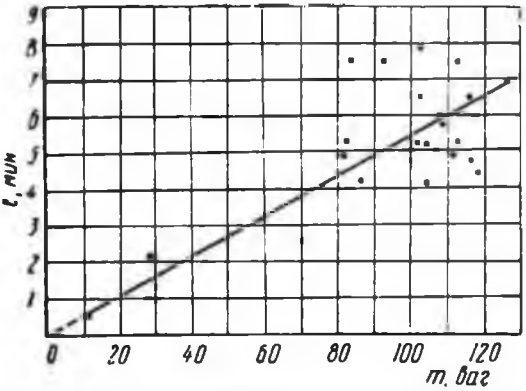

Рис. 5.

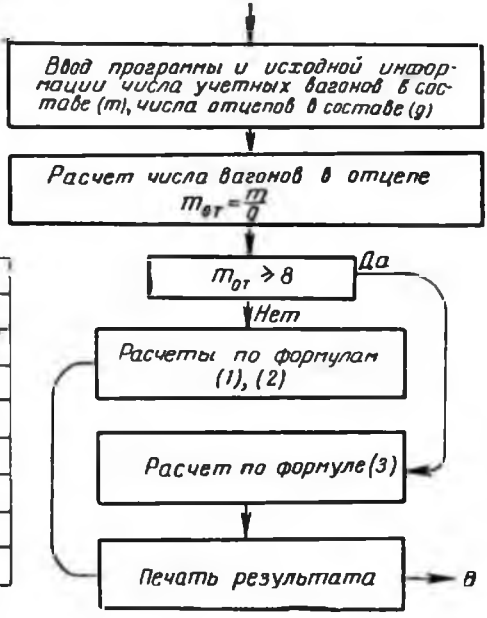

Рис. 6.

лено, что при среднем числе учетных вагонов в отцепах более семи может быть принята линейная зависимость продолжительности роспуска составов от общего числа вагонов в составе m (рис. 5).

Поскольку прямая проходит через начало системы координат, эта зависимость может быть выражена уравнением $t_p = am$. Значение коэффициента a определяется из графика (рис. 5).

$$t_p = 0{,}055 \cdot m \quad (3)$$

Полученные аналитические выражения зависимостей продолжительности роспуска составов от различных факторов могут быть положены в основу алгоритма определения продолжительности роспуска составов на сортировочных горках (рис. 6).