

Федеральное государственное бюджетное образовательное учреждение
высшего образования
«Уфимский государственный нефтяной технический университет»
Кафедра «Технологические машины и оборудование»

Отчет
о практической работе
«Математическая обработка результатов наблюдений
при экспоненциальном законе распределения»

Студент гр. БМР 19-01

А.А. Петров

Доцент каф. ТМО

А.Х. Габбасова

Уфа 2021

Цель. Для возможности прогнозирования надежности объекта выбрать закон распределения при заданных значениях наработки до отказа ряда аналогичных объектов.

Покажем обработку результатов наблюдений для определения оценок числовых характеристик и вида закона распределения случайных величин x_i на примере наработки на отказ насоса, перекачивающего горячую кислоту.

В результате наблюдений получено сто случайных значений ($n = 100$) времени безотказной работы насоса $x_i = t_i$ (ч), которые приведены таблице 1.

Таблица 1 - Случайные значения времени безотказной работы насоса t , ч

314	8	77	133	35	31	68	32	25	169
249	216	16	17	367	48	18	62	54	49
145	6	95	29	224	6	47	70	16	105
107	26	12	15	128	108	7	125	7	78
85	8	107	19	206	27	8	144	12	101
56	6	225	18	146	43	209	55	70	18
34	49	16	64	79	38	46	4	187	25
22	165	28	294	17	61	5	14	78	27
17	9	32	21	41	4	18	34	154	6
8	370	1	21	63	28	149	368	189	25

Приведенный в таблице 1 экспериментальный статистический материал для придания ему наглядности и компактности целесообразно представить в виде **статистического (вариационного) ряда** – по возрастанию.

В множестве данных находится минимальный член ряда - 1 ч и максимальный - 370 ч.

Размах ряда составляет, ч

$$t_{max} - t_{min} = 370 - 1 = 369.$$

Весь диапазон значений случайной величины t_i ($n = 100$) разбивается на интервалы.

Для удобства расчетов интервалы целесообразно принимать равными.

Примерная величина интервала Δt определяется по формуле

$$\Delta t = \frac{t_{max} - t_{min}}{1 + 2,31 \cdot \lg n} \quad (1)$$

$$\Delta t = \frac{370 - 1}{1 + 2,31 \cdot \lg 100} = 66.$$

Если при выбранных по формуле (1) равных интервалах количество значений случайной величины в интервале оказывается меньше 10 (таблица 2, столбец 4), то принимаются интервалы различной длины (таблица 2, столбец 2, столбец 3).

Количество интервалов рекомендуется брать от 7 до 15. Большое число интервалов принимается для весьма обширного и довольно однородного статистического материала.

Число интервалов статистического распределения в примере $k = 10$. Для каждого интервала проведем подсчеты и представим их в таблице 2.

Таблица 2 - Обработка статистического ряда

№	Интервал	Размах Δt_i	n_i	$P_i = n_i / n$	$t_{i(серед.)}$	$P_i t_i$	$F^*(t) = \Sigma(n_i / n)$	$f^*(t) = n_i / (n\Delta t)$	$P_i [t_i - M(t)]^2$
1	2	3	4	5	6	7	8	9	10
1	0-10	10	15	0,15	5	0,75	0,15	0,0150	735,00
2	10-20	10	14	0,14	15	2,10	0,29	0,0140	504,00
3	20-30	10	11	0,11	25	2,75	0,40	0,0110	275,00
4	30-50	20	15	0,15	40	6,00	0,55	0,0075	183,75
5	50-80	30	14	0,14	65	9,10	0,69	0,0047	14,00
6	80-110	30	7	0,07	95	6,65	0,76	0,0023	28,00
7	110-150	40	8	0,08	130	10,40	0,84	0,0020	242,00
8	150-190	40	5	0,05	170	8,50	0,89	0,0013	451,25
9	190-250	60	6	0,06	220	13,20	0,95	0,0010	1261,50
10	250-370	120	5	0,05	310	15,50	1,00	0,0004	2761,25
Σ	-	-	100	1,00	-	$M(t) =$ $= 75 \text{ ч}$	-	-	$D(t) =$ $= 6456 \text{ ч}^2$

В таблице 2 приведены следующие результаты расчетов:

В колонке 1 даны номера разрядов разбивки интервала варьирования.

В колонке 2 - границы каждого интервала, причем условимся, что каждый предыдущий интервал содержит конечную точку, а каждый последующий не содержит точки начала.

В колонке 3 показана числовая величина размаха интервала Δt .

В колонке 4 - количество значений случайной величины (из таблицы 1), попавших в интервал, т. е. **абсолютная частота n_i** .

В колонке 5 подсчитана **относительная частота (часть)**, или эмпирическая вероятность

$$P_i = n_i / n, \quad (2)$$

где n_i / n - накопленная относительная частота всех интервалов должна быть равна единице, что служит проверкой правильности вычисления частоты для каждого интервала. В примере $n = 100$.

В колонке 6 отмечены координаты t_i (серед.) - середины каждого интервала из колонки 2.

В колонке 7 приведены произведения значений из колонок 5 и 6 ($P_i \cdot t_i$), которые в сумме дают координату центра распределения, т.е. **статистическое среднее (математическое ожидание) $M(t)$**

$$\Sigma(P_i \cdot t_i) = M(t) = 75 \text{ (ч)}.$$

В колонке 8 дана **накопленная частота $\Sigma(n_i / n)$** , или функция распределения **$F^*(t)$** . Определяется по данным колонки 5 (складываются значения текущего и предыдущих интервалов).

В колонке 9 - **эмпирическая плотность вероятности $n_i / (n\Delta t)$** , или **$f^*(t)$** . Определяется по данным колонки 3 и колонки 5.

В колонке 10 указано произведение $P_i [t_i - M(t)]^2$, служащее для определения **статистической дисперсии $D(t)$**

$$D(t) = \Sigma (P_i [t_i - M(t)]^2) = \Sigma (P_i [t_i - 75]^2) = 6456 \text{ (ч}^2\text{)}.$$

Среднеквадратическим отклонением $\sigma(t)$ будет положительное значение корня квадратного из дисперсии

$$\sigma(t) = \sqrt{D(t)} = \sqrt{6456} = 80 \text{ (ч)}. \quad (3)$$

Коэффициент вариации определяется как

$$V(t) = \frac{\sigma(t)}{M(t)} = \frac{80}{75} = 1,07. \quad (4)$$

Часто в статистических исследованиях используют следующие характеристики:

- **среднеквадратическая ошибка определения среднего арифметического (среднеквадратическая ошибка определения математического ожидания)**

$$\Delta M(t) = \frac{\sigma(t)}{\sqrt{n}} = \frac{80}{\sqrt{100}} = 8 \text{ (ч)}; \quad (5)$$

- **среднеквадратическая ошибка определения среднеквадратического отклонения**

$$\Delta \sigma(t) = \frac{\sigma(t)}{\sqrt{2n}} = \frac{80}{\sqrt{200}} = 5,7 \text{ (ч)}. \quad (6)$$

Тогда, округляя значения отклонений, получим

$$M(t) = 75 \pm 8 \text{ (ч)}.$$

На рисунке 1 для построения гистограммы по оси абсцисс t откладываются интервалы Δt_i (см. таблицу 2, колонка 2) случайной величины t_i и на каждом из интервалов строится прямоугольник с площадью, равной частоте появления случайной величины в данном интервале.

Высоты прямоугольников пропорциональны соответствующим частотам и равны эмпирической плотности вероятности $[n_i / (n\Delta t_i)]$ (см. таблицу 2, данные колонки 9) для каждого интервала. На рисунке 1 представлена точечная гистограмма распределения, где значения статистической вероятности отложены от середины каждого интервала.

Исходя из характера гистограммы, можно **предположить**, что исследуемая **случайная величина распределена по экспоненциальному закону**. Об этом свидетельствует также почти полное совпадение по величине математического ожидания $M(t) = 75$ ч и среднеквадратического отклонения $\sigma(t) = 80$ ч случайной величины t (коэффициент вариации $V \approx 1$).

Выравнивающий график функции $f(t)$ строим по данным таблицы 3 (колонка 5).

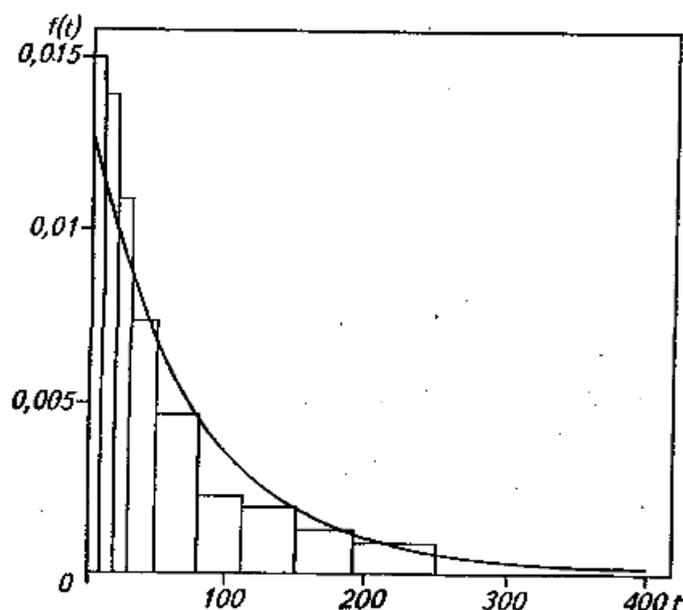


Рисунок 1 – Гистограмма наработки на отказ и выравнивающая кривая

Приняв в качестве математического ожидания наработки на отказ его оценку (статистическое среднее) $M(t) = 75$ ч, можно записать

$$f(t) = \lambda \cdot e^{-\lambda t}, \text{ или } f(t) = \lambda \cdot \exp(-\lambda \cdot t_i). \quad (7)$$

Поскольку при экспоненциальном законе распределения $\lambda = 1 / M(t) = \text{const}$, то

$$f(t) = \lambda \cdot \exp(-\lambda \cdot t_i) = 0,013 \cdot e^{-0,013 \cdot t}, \quad (8)$$

где $\lambda = 1 / M(t) = 1 / 75 = 0,013$.

В таблице 3 приведены следующие результаты расчетов:

В колонке 2 значение $t_{i(гран.)}$ границы интервалов из колонки 2 таблицы 2.

В колонке 3 значения $\lambda \cdot t_i$.

В колонке 4 функции ($e^{-\lambda \cdot t}$) можно определить, используя таблицу e^{-x} приложения А.

Таблица 3 - Теоретические значения вероятностей

№	$t_{i(гран.)}$	$\lambda \cdot t_i$	$e^{-0,013 \cdot t}$	$f(t)$	$P'(t_i)$	$\frac{n \cdot P'(t_i)}{P'(t_i)}$	n_i	$\frac{[n_i - nP'(t_i)]^2}{nP'(t_i)}$	$F(t)$
1	2	3	4	5	6	7	8	9	10
1	0	0	1	0,0130	-	-	-	-	0
2	10	0,13	0,878	0,0114	0,122	12,2	15	0,643	0,122
3	20	0,26	0,770	0,0100	0,108	10,8	14	0,948	0,230
4	30	0,39	0,677	0,0088	0,093	9,3	11	0,311	0,323
5	50	0,65	0,522	0,0068	0,155	15,5	15	0,016	0,478
6	80	1,04	0,353	0,0046	0,169	16,9	14	0,500	0,647
7	110	1,43	0,239	0,0031	0,114	11,4	7	1,700	0,761
8	150	1,95	0,142	0,0018	0,097	9,7	8	0,298	0,858
9	190	2,47	0,085	0,0011	0,037	3,7	5	0,457	0,915
10	250	3,25	0,039	0,0005	0,046	4,6	6	0,426	0,961
11	370	4,81	0,008	0,0001	0,031	3,1	5	1,160	0,992
							100	$\Sigma = 6,463$	

В колонке 5 приведены результаты расчета значений **плотности вероятности $f(t)$** на границах интервалов, полученные по формуле (8) с использованием функции ($e^{-\lambda \cdot t}$)_i (таблица 3, колонка 4).

На гистограмме (рисунок 1) построена **выравнивающая кривая распределения**, представляющая собой график функции $f(t)$, которая, сохраняя в основном существенные особенности статистического распределения, свободна от случайных неправильностей хода гистограммы.

При подборе теоретической кривой распределения между ней и статистическим распределением неизбежны некоторые расхождения. Они могут объясняться случайными обстоятельствами, связанными с ограниченным числом опытных данных, или являться существенными - связанными с тем, что подобранная кривая плохо выравнивает данные распределения.

Установить это можно с помощью **критерия согласия Пирсона χ^2**

$$\chi^2 = \sum_{i=1}^k \frac{[n_i - nP'(t_i)]^2}{nP'(t_i)}, \quad (9)$$

где k - число интервалов статистического распределения, в примере $k = 10$;
 n_i - количество значений случайной величины в каждом интервале (см. таблицу 2, колонка 4);
 n - общее число наблюдаемых значений случайной величины, в примере $n = 100$;
 $P'(t_i)$ - теоретическая вероятность попадания случайной величины в i -й интервал (таблица 3, колонка 6).

В колонке 6 таблицы 3 приведены значения **вероятностей попадания случайной величины в i -й интервал $P'(t_i)$** . Они численно равны приращению функции распределения на интервале (от предыдущих значений колонки 4 вычитаем текущее значение)

$$P'(t_i < t < t_{i+1}) = F(t_i + 1) - F(t_i) \quad (10)$$

или

$$P'(t_i) = [(e^{-\lambda \cdot t})_{i-1}] - [(e^{-\lambda \cdot t})_i].$$

В колонке 7 рассчитаны значения $n \cdot P'(t_i)$, где $n = 100$.

В колонке 8 - количество значений случайной величины, попавших в интервал, т. е. **абсолютная частота n_i** (из колонки 4 таблицы 2).

В колонке 9, так как для интервала 0-10 абсолютная частота $n_i = 15$ (см. таблицу 3, колонка 4), а значение $nP'(t_i) = 12,2$ (см. таблицу 3, колонка 7), то распределение

$$\chi^2 = [n_i - nP'(t_i)]^2 / nP'(t_i) = (15 - 12,2)^2 / 12,2 = 0,643.$$

и так далее проводим расчет в колонке 9.

Распределение χ^2 зависит от параметра R , называемого числом «степеней свободы». **Число «степеней свободы» R** равно числу интервалов k за вычетом числа независимых условий (связей) S , наложенных на частоты n_i / n

$$R = k - S. \quad (11)$$

Число связей S для экспоненциального закона распределения случайной величины $S = 2$, для нормального $S = 3$.

По специальной таблице (приложение Б), можно для полученного значения χ^2 и определенного числа «степеней свободы» R найти P - вероятность того, что величина, распределенная по закону χ^2 , превзойдет это значение.

При этом, если получаемая вероятность P больше 0,3 - 0,4, обычно признают, что экспериментальные данные не противоречат принятому теоретическому закону распределения случайной величины.

В примере $\chi^2 = 6,463$ (колонка 9 таблицы 3) и **число степеней свободы $R = 10 - 2 = 8$** .

По таблице в приложении Б для значений $\chi^2 = 6,463$ и $R = 8$ **находим $P=0,6$** .

К наиболее часто употребляемым критериям согласия наряду с критерием «хи-квадрат» относится также **критерий Колмогорова D** .

В качестве меры расхождения между теоретическим и статистическим распределениями Колмогоров рассматривает максимальное значение модуля разности между статистической функцией распределения $F^*(t)$ и соответствующей теоретической функцией распределения $F(t)$

$$D = \max |F^*(t) - F(t)|. \quad (12)$$

Величина D находится из графиков $F^*(t)$ и $F(t)$.

На рисунке 2 приведены **графики статистической и теоретической (экспоненциальной) функций распределения** для рассматриваемого примера.

Для построения **интегральной статистической функции распределения $F^*(t)$** используется накопленная частота $\sum n_i / n$ (см. таблицу 2, колонка 8).

В колонке 10 таблицы 3 для построения **теоретической функции распределения $F(t)$** (экспоненциальной) воспользуемся выражением для этой функции по данным таблицы 3, колонки 4

$$F(t) = 1 - e^{-\lambda t} = 1 - e^{-0,013 \cdot t}.$$

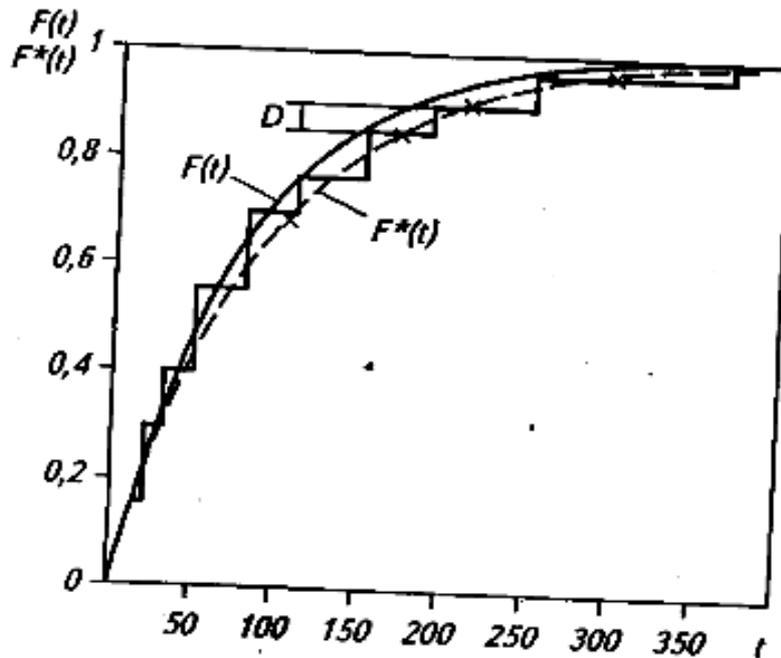


Рисунок 2 - Графики статистической $F^*(t)$ и теоретической $F(t)$ функций распределения

Из рисунка 2 и из сравнения данных колонки 8 таблицы 2 и колонки 10 таблицы 3:

$$\begin{aligned}
 D &= |F^*(10) - F(10)| = |0,15 - 0,122| = 0,028; \\
 D &= |F^*(20) - F(20)| = |0,29 - 0,230| = 0,060; \\
 D &= |F^*(30) - F(30)| = |0,40 - 0,323| = 0,077 = D_{max}; \\
 D &= |F^*(50) - F(50)| = |0,55 - 0,478| = 0,072; \\
 D &= |F^*(80) - F(80)| = |0,69 - 0,647| = 0,043; \\
 D &= |F^*(110) - F(110)| = |0,76 - 0,761| = 0,001; \\
 D &= |F^*(150) - F(150)| = |0,84 - 0,858| = 0,018; \\
 D &= |F^*(190) - F(190)| = |0,89 - 0,915| = 0,025; \\
 D &= |F^*(250) - F(250)| = |0,95 - 0,961| = 0,011; \\
 D &= |F^*(370) - F(370)| = |1,00 - 0,992| = 0,008, -
 \end{aligned}$$

видно, что **максимальная разница** значений $F^*(t_i)$ и $F(t_i)$ наблюдается **при $t_i=30$ ч.** При этом критерий Колмагорова D_{max} составляет **0,077**.

Далее определяется величина

$$\lambda^* = D\sqrt{n},$$

$$\lambda^* = 0,077 \cdot \sqrt{100} = 0,77.$$

По приложению В находится **вероятность $P(\lambda^*)$**

$$P(\lambda^*) = P(0,77) = 0,544.$$

Вывод. Вероятность $P(\lambda^*) = 0,544$ не является малой, таким образом, гипотеза об экспоненциальном законе распределения времени безотказной работы насоса подтверждается также и критерием Колмогорова.