Практическое занятие 2

Тема: **Статистическая обработка массива однородных величин**

По статистическим данным определяем $x_{ср}$ – среднее значение и S – среднее квадратичное отклонение для всего массива статистических данных рассчитываются по общим формулам:

$$x_{cp} = \frac{\sum_{i=1}^{N} x_i}{N} \quad ; \qquad S = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_i - x_{cp})^2} \qquad (4)$$

На основании условий и признаков, при которых следует ожидать появление того или иного закона распределения, а также по виду графика эмпирической функции распределения, производится предварительный выбор закона распределения исследуемой случайной величины.

Наиболее распространенные в теории надежности законы распределения случайных величин

- нормальный закон распределения

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \cdot \exp\left(-\frac{(x-a)^2}{2\sigma^2}\right)$$

- закон распределения Вейбулла

$$f(x) = \lambda \cdot m \cdot x^{m-1} \cdot \exp(-\lambda \cdot x^m)$$

- экспоненциальный закон распределения

$$f(x) = \lambda \cdot \exp(-\lambda \cdot x)$$

В зависимости от принятого закона распределения случайных величин производиться расчет его параметров:

- для двухпараметрического закона нормального распределения $a = x_{ср}$, $\sigma = S$;
- для двухпараметрического закона распределения Вейбулла по формуле

$$V = \frac{S}{x_{cp}} = \sqrt{\frac{\Gamma(1+\frac{2}{m})}{\Gamma^2(1+\frac{1}{m})^{-1}}} \qquad (5)$$

определяется параметр *V* и затем по таблице 1 находится величина *m*.

В случае отсутствия соответствующего значения в таблице задача может быть решена подбором значения *m*, преобразующего уравнение (5) в тождество.

Второй параметр λ определяется из уравнения:

$$\lambda = \left(\frac{\Gamma\left(1+\frac{1}{m}\right)}{x_{cp}} \right)$$

$$\lambda = 1 / x_{cp}$$

- для однопараметрического экспоненциального закона

Таблица 1.

Зависимость между параметрами *V* и *m* для распределения Вейбулла

V	***m***	***V***	***m***	***V***	***m***	***V***	***m***
0,36	**3,03**	**0,58**	**1,78**	**0,80**	**1,26**	**1,50**	**0,68**
0,38	**2,85**	**0,60**	**1,72**	**0,85**	**1,18**	**1,60**	**0,64**
0,40	**2,70**	**0,62**	**1,66**	**0,90**	**1,11**	**1,80**	**0,59**
0,42	**2,55**	**0,64**	**1,60**	**0,95**	**1,05**	**2,00**	**0,54**
0,44	**2,42**	**0,66**	**1,54**	**1,00**	**1,00**	**2,20**	**0,50**
0,46	**2,31**	**0,68**	**1,49**	**1,05**	**0,95**	**2,40**	**0,47**
0,48	**2,20**	**0,70**	**1,45**	**1,10**	**0,91**	**2,60**	**0,45**
0,50	**2,10**	**0,72**	**1,41**	**1,15**	**0,87**	**3,00**	**0,41**
0,52	**2,01**	**0,74**	**1,37**	**1,20**	**0,84**	**3,50**	**0,37**
0,54	**1,93**	**0,76**	**1,33**	**1,30**	**0,78**	**4,00**	**0,35**
0,56	**1,85**	**0,78**	**1,29**	**1,40**	**0,73**	**5,00**	**0,31**

Таблица 2. **Значения Гамма-функции**

x	*Г(x)*	*x*	*Г(x)*	*x*	*Г(x)*	*x*	*Г(x)*
1 00	1,0000	1,25	0,9064	1,50	0,8862	1,75	0,9191
1,01	0,9943	1,26	0,9044	1,51	0,3366	1,76	0,9214
1,02	0,9888	1,27	0,9025	1,52	0,8870	1,77	0,9238
1,03	0,9836	1,28	0,9007	1,53	0,8376	1,78	0,9262
1,04	0,9784	1,29	0,8990	1,54	0,8882	1,79	0,9288
1,05	0,9735	1,30	0,8975	1,55	0,8889	1,80	0,9314
1,06	0,9687	1,31	0,8960	1,56	0,8896	1,81	0,9341
1,07	0,9642	1,32	0,8946	1,57	0,8905	1,32	0,9369
1,03	0,9597	1,33	0,8934	1,58	0,8914	1,33	0,9397
1,09	0,9555	1,34	0,8922	1,59	0,8924	1,34	0,9426
1,10	0,9514	1,35	0,8912	1,80	0,8935	1,35	0,9456
1,11	0,9474	1,36	0,8902	1,61	0,8947	1,86	0,9487
1,12	0,9436	1,37	0,8893	1,62	0,8959	1,87	0,9518
1,13	0,9399	1,38	0,8885	1,63	0,8972	1,88	0,9551
1 14	0,9364	1,39	0,8878	1,64	0,8986	1,39	0,9534
1 15	0,9330	1,40	0,8873	1,65	0,9001	1,90	0,9613
1,16	0,9298	1,41	0,8868	1,66	0,9017	1,91	0,9652
1,17	0,9267	1,42	0,8864	1 67	0,9033	1,92	0,9688
1,18	0,9237	1,43	0,8860	1,68	0,9050	1,93	0 9724
1,19	0,9209	1,44	0,8858	1,69	0,9063	1,94	0,9761
1,20	0,9182	1 45	0,8857	1,70	0,9086	1,95	0,9799
121	0,9156	1,46	0,8856	1,71	0,9106	1,96	0,9837
1,22	0,9131	1,47	0,8856	1,72	0,9126	1,97	0,9877
1,23	0,9108	1,48	0,8858	1,73	0,9147	1,98	0,9917
1,24	0,9085	1,49	0,8860	1,74	0,9168	1,99	0,9958
1,25	0,9064	1,50	0,8862	1,75	0,9191	2,00	1,0000

Проверка соответствия выбранного закона распределения случайных величин имеющимся статистическим данным

Для проверки соответствия выбранного закона распределения случайных величин имеющимся статистическим данным необходимо произвести проверку графическим способом, по критерию Колмогорова или по критерию χ_2 (хи – квадрат).

Во всех случаях для проверки необходимо определить теоретические значения количества отказов для каждого из интервалов.

Расчет производиться по формуле

$$\Delta n_i^{'} = a\left(x_p\right) \cdot \Delta x \cdot N$$

Расчет производиться для середины соответствующего интервала

$$x_p = t_{i\ cp}$$

Проверка соответствия выбранного закона распределения случайных величин имеющимся статистическим данным

Наиболее прост и нагляден графический способ проверки.

На графике изменения частостей (гистограмме или полигоне распределения), построенном по статистическим данным, строиться аналогичный график по полученным теоретическим данным. Совпадение графиков свидетельствует о правильности выбора закона распределения.

Проверка **по критерию Колмогорова** производиться в следующей последовательности:

1. Для всех интервалов определяются значения модуля разности между эмпирическими и теоретическими накопленными частотами по формуле

$$d_i = \left| n_i - n_i' \right|$$

2. Определяется максимальное значение di max и величина λк:

$$\lambda_к = \frac{d_{i\max}}{\sqrt{N'}}$$, где $$N' = \sum_{i=1}^{k} \Delta n_i'$$

Проверка соответствия выбранного закона распределения случайных величин имеющимся статистическим данным

По таблице 3 определяется вероятность $P(\lambda_k)$ совпадения статистических и теоретических данных. При значении $P(\lambda_к)$, близком к 1, правильность выбранного закона распределения считается подтвержденной.

Таблица 3

Значения *$P(\lambda_к)$* критерия Колмогорова

$\lambda_к$	$P(\lambda_к)$	$\lambda_к$	$P(\lambda_к)$	$\lambda_к$	$P(\lambda_к)$	$\lambda_к$	$P(\lambda_к)$
0,30	1,000	0,64	0,8073	1,00	0,2700	1,30	0,0032
0,35	0,9997	0,65	0,7920	1,10	0,1777	1,90	0,0015
0,40	0,9972	0,70	0,7112	1,20	0,1122	2,00	0,0007
0,45	0,9874	0,75	0,6272	1,30	0,0631	2,10	0,0003
0,50	0,9639	0,80	0,5441	1,40	0,0397	2,20	0,0001
0,55	0,9228	0,85	0,4653	1,50	0,0222	2,30	0,0001
0,53	0,8896	0,90	0,3927	1,60	0,0120	2,40	0,0000
0,50	0,8643	0,95	0,3275	1,70	0,0062	2,50	0,0000

Проверка соответствия выбранного закона распределения случайных величин имеющимся статистическим данным

Критерий согласия хи-квадрат применяется при большом количестве опытных данных и в случаях, если теоретические значения параметров функции распределения неизвестны. Сравнение статистических данных с теоретическими значениями функции распределения производиться следующим образом:

1. Производиться группировка (объединение) интервалов таким образом, чтобы в каждом из интервалов было бы не менее 4-6 отказов.

2. Определяется мера расхождения *χ2* по формуле

$$\chi^2 = \sum_{i=1}^{r} \frac{(n_i - n_i')^2}{n_i'}$$

Определяется число степеней свободы k по формуле

$k = r - m - 1$,

где r – количество интервалов после группировки;

m – количество параметров в законе распределения (для распределения Вейбулла - λ и m).

Проверка соответствия выбранного закона распределения случайных величин имеющимся статическим данным

3. По найденным значениям *χ2* и *k* по таблице 4 находиться вероятность Р того, что величина, имеющая распределение хи – квадрат с *k* степенями свободы, превзойдет полученное значение *χ2*.

Если эта вероятность мала (*0,1-0,2*), то принятое теоретическое распределение должно быть отвергнуто как неправдоподобное. Большие значения вероятности (больше *0,7-0,8*) свидетельствуют, что допущены какие-то систематические ошибки или погрешности при предварительной обработке статистических данных.

Таким образом, при $0,2 \leq P \leq 0,7$ считается, что принятая гипотеза распределения отказов, определенных по принятой теории распределения, соответствует статистическим данным.

Проверка соответствия выбранного закона распределения случайных величин имеющимся статическим данным

Таблица 4. **Значения χ^2 в зависимости от k и P**

	P													
k	***0,99***	***0,93***	***0,95***	***0,90***	***0,80***	***0,70***	***0,50***	***0,30***	***0,20***	***0,10***	***0,05***	***0,02***	***0,01***	***0,00***
1	0,000	0,001	0,004	0,016	0,064	0,148	0,455	1,074	1,642	2,71	3,84	6,41	6,61	10,35
2	0,020	0,040	0,103	0,211	0,445	0,713	1,335	2,41	3,22	4,50	5,99	7,32	9,21	13,32
3	0,115	0,135	0,352	0,534	1,005	1,424	2,37	3,65	4,64	5,25	7,32	9,34	11,34	15,27
4	0,297	0,429	0,711	1,064	1,649	2,20	3,36	4,88	5,99	7,78	9,49	11,57	13,28	13,46
5	0,554	0,752	1,145	1,610	2,34	3,00	4,35	6,06	7,29	9,24	11,07	13,39	15,09	20,5
6	0,872	1,134	1,535	2,20	3,07	3,83	5,35	7,23	8,56	10,64	12,59	15,03	16,81	22,5
7	1,239	1,564	2,17	2,33	3,82	4,67	6,35	8,38	9,80	12,02	14,07	16,52	13,43	24,3
8	1,646	2,03	2,73	3,49	4,59	5,53	7,34	9,52	11,03	13,36	15,51	18,17	20,1	26,1
9	2,09	2,53	3,32	4,17	5,38	6,39	8,34	10,66	12,24	14,68	16,92	19,68	21,7	27,9
10	2,56	3,06	3,94	4,86	6,18	7,27	9,34	11,78	13,44	15,99	13,81	21,2	23,2	29,6
11	3,05	3,61	4,58	5,58	6,99	8,15	10,34	12,90	14,63	17,28	19,68	22,6	24,7	31,3
12	3,57	4,18	5,23	6,30	7,81	9,03	11,34	14,01	15,81	18,55	21,0	24,1	25,2	32,9

Пример (продолжение)

Задание: Произвести статистическую обработку массива однородных экс- периментальных данных об отказах УЭЦН.

Исходные данные (наработка, сут):

340 774 279 517 470 934 76 397

570 221 679 983 756 701 277 498

824 432 386 506 597 641 209 214

272 384 411 729 689 682 540 650

561 716 404 575 88 189 124 721

4. Определяем границы всех интервалов, средние значения интервалов, значения записываются в табл.1 (графы 2 ,3).

5. Определяем общее количество отказов N = 40, количества отказов для всех интервалов n и частости m, значения записываются в табл.1 (графы 4,5).

№ интер.	Границы интервалов	$t_{ср\ инт}$	n	m	n'	d
1	*70 - 161*	*116*	*3*	*0.075*	*1.86*	*1.14*
2	*162 - 253*	*208*	*4*	*0.1*	*3.70*	*1.44*
3	*254 - 345*	*300*	*4*	*0.1*	*5.21*	*0.22*
4	*346 - 437*	*392*	*6*	*0.15*	*6.05*	*0.17*
5	*438 - 529*	*484*	*4*	*0.1*	*6.08*	*1.91*
6	*530 - 621*	*576*	*5*	*0.125*	*5.4*	*2.31*
7	*622 - 713*	*668*	*6*	*0.15*	*4.28*	*0.59*
8	*714 - 805*	*760*	*5*	*0.125*	*3.05*	*1.36*
9	*806 - 897*	*852*	*1*	*0.025*	*1.95*	*0.41*
10	*898 - 990*	*944*	*2*	*0.05*	*1.12*	*1.29*

6. По имеющимся статистическим данным строим гистограмму и лома- ную - полигон частот отказов (рис.1).

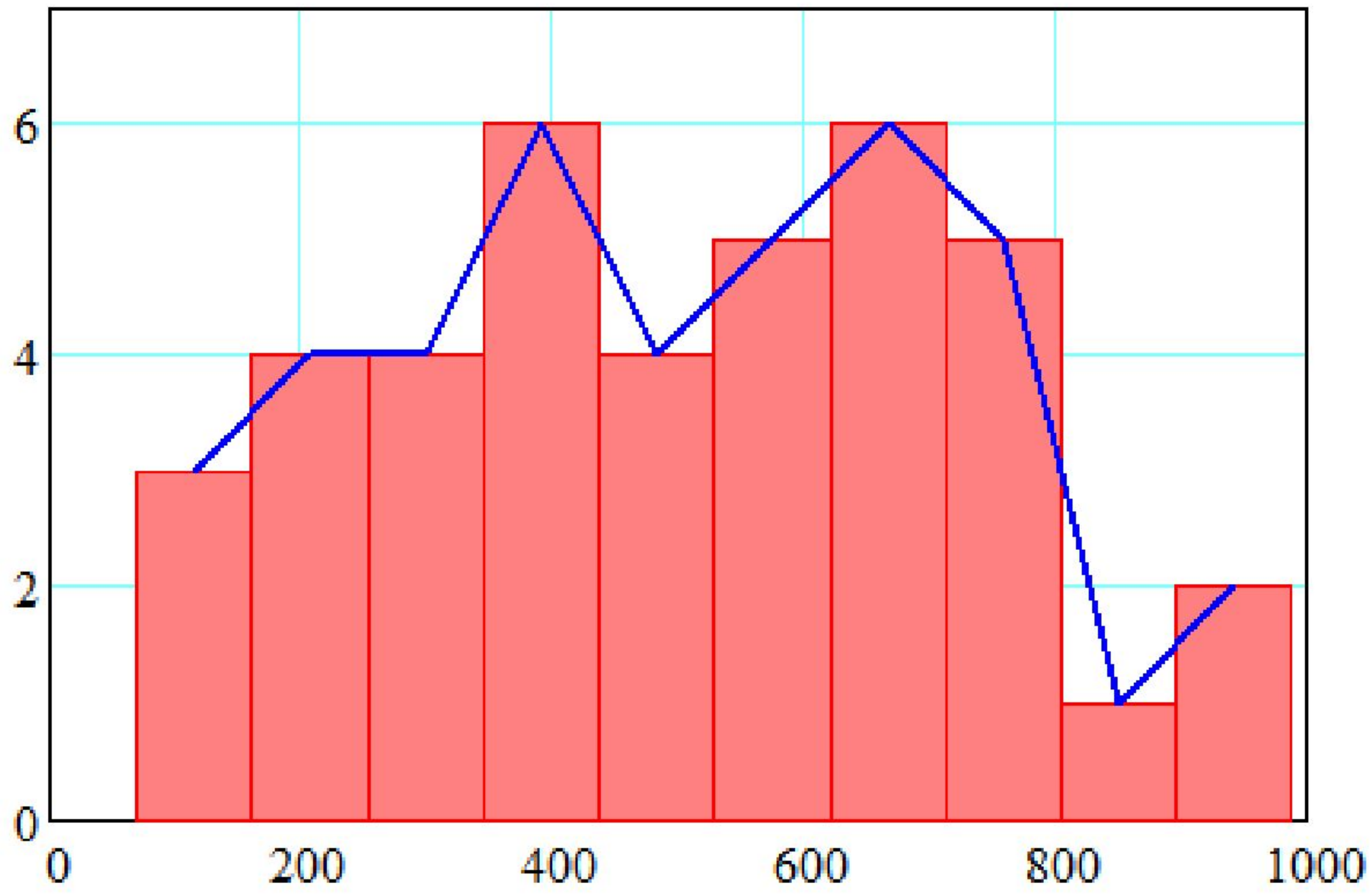


Рис.1 Гистограмма (красным) и полигон частот (синим)

7. Определяем величину среднего значения

$$t_{cp} = \frac{\sum t_{cpi} n_i}{N}, \qquad t_{cp} = 500{,}4 \text{ сут.}$$

8. Определяем величины среднего значения среднеквадратичного отклонения S:

$$S = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(t_i - t_{cp}\right)^2}, \qquad S = 228{,}7 \text{ сут.}$$

9. Исходя из рис. 1 выбираем закон распределения Вейбулла

$$f(x) = \lambda \cdot m \cdot x^{m-1} \cdot \exp(-\lambda \cdot x^m)$$

и определяем его параметр.

Вводим вспомогательный параметр V, численное значение которого находим через

t_{cp} и S.

$$V = S / t_{cp} = 0{,}457$$

По таблице зависимости между параметрами V и m распределения Вейбулла определяем $m = 2{,}3$.

Находим значение λ:

$$\lambda = \left(\frac{\Gamma\left(1+\frac{1}{m}\right)}{t_{cp}} \right)^{m} = \left(\frac{\Gamma\left(1+\frac{1}{2,3}\right)}{t_{cp}} \right)^{2,3} = \left(\frac{0,8859}{500,4} \right)^{2,3} = 4,68 \cdot 10^{-7}$$

10. Определяем теоретические значения функции плотности распределения отказов по формуле

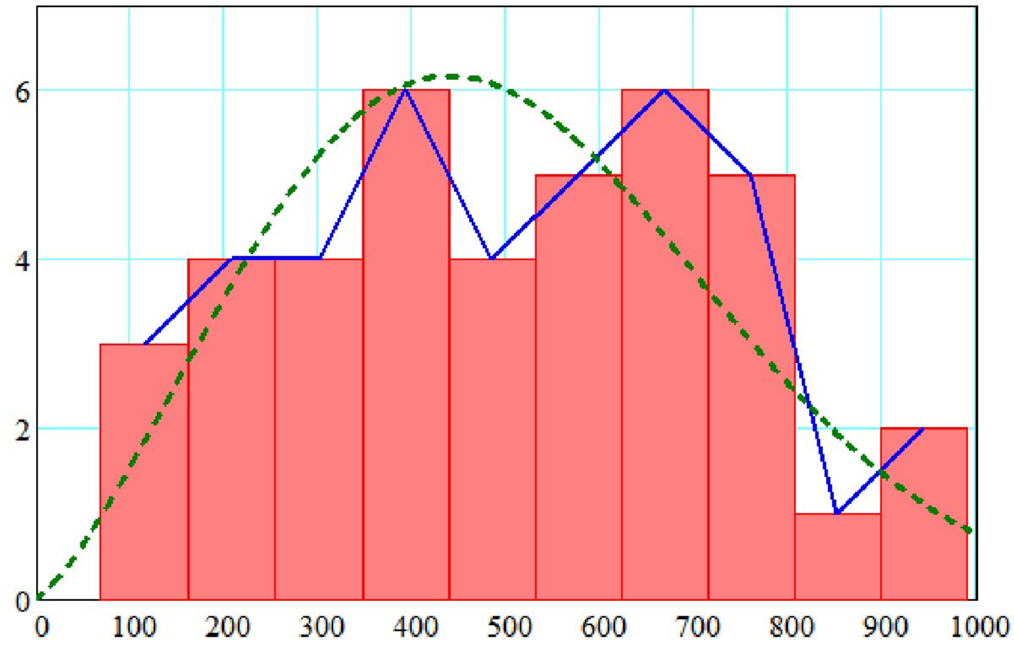
$$n_i' = f\left(t_{cp\_uнт}\right) \cdot \Delta t \cdot N$$

Полученные значения количества отказов записываем в таблицу 1 (6 графа).

11. Производим проверку соответствия распределения статистических данных принятому закону распределения.

а) Графическим способом.

Получено удовлетворительное соответствие теоретической криво и ломаной полигона частот.



б) По критерию Колмогорова. Определяем накопленное количество отказов для каждого интервала по статистическим и теоретическим данным, а также d-модули их разности.

Результаты записываются в табл.1 (графа 7).

$$d_i = \left| n_i - n_i' \right|$$

$$N' = \sum_{i=1}^{k} n_i'$$

$$\lambda_k = \frac{d_{i\max}}{\sqrt{N'}} = 0,371$$

По справочникам определяется вероятность $P(\lambda_k) = 0.9985$.

Полученное значение близко к единице, что подтверждает правильность выбранного теоретического закона распределения.

в) По критерию ***χ2*** (хи-квадрат).

Определяем количества отказов в интервалах отказов по экспериментальным и теоретическим данным.

Находим

$$\chi^2 = \sum_{i=1}^{r} \frac{(n_i - n_i')^2}{n_i'}$$

χ2 = 4.83

Определяем число степеней свободы *k* по формуле *k = r – m – 1*,

где *r* = 7 - количество интервалов,

m = 2 - количество параметров закона распределения,

k = 4

По справочникам находим вероятность P = 0,3 , это свидетельствует о том, что принятый закон распределения соответствует статистическим данным.