Компьютерные методы моделирования строительства скважин.

Лабораторная работа № 2.

ПРОВЕРКА СООТВЕТСТВИЯ ВЫБОРКИ НОРМАЛЬНОМУ ЗАКОНУ РАСПРЕДЕЛЕНИЯ

Цель работы: овладение студентом способами построения эмпирической и теоретической (нормальной) кривой распределения; выработка умения и навыков применения критериев согласия для проверки выдвинутой статистической гипотезы.

Содержание работы: на основе дискретного вариационного ряда, полученного в *лабораторной работе № 1*, выполнить следующее:

1. Построить эмпирическую (полигон) и теоретическую (нормальную) кривую распределения.

2. Проверить согласованность эмпирического распределения с теоретическим нормальным, применяя критерии:

а) критерий Пирсона;

б) критерий Романовского

в) один из критериев: Колмогорова(чётный вариант), Ястремского; (нечётный)

г) приближенный критерий.

Сделать вывод по результатам работы.

Методика выполнения работы

1. За основу взять дискретный вариационный ряд, полученный в ***лаб.1*** квадратичное отклонение и среднее арифметическое для него $\bar{x}, S$.

В одной координатной плоскости построить эмпирическую и теоретическую кривые.

Эмпирическая кривая распределения представляет собой полигон частот (см. лабораторную работу № 1). Полигон строить в координатах (x_i, n_i), с помощью ломаной.

Для построения теоретической (нормальной) кривой найти координаты точек $\left(x_i, n_i'\right)$, для чего рассчитать теоретические частоты n_i', пользуясь таблицей. Значения $\varphi(u_i)$ найти по приложению 1.

x_i	n_i	$x_i - \bar{x}$	$u_i = \frac{x_i - \bar{x}}{S}$	$\varphi(u_i)$	$y_i = \frac{nh}{S}\varphi(u_i)$	n_i'
⋮	⋮	⋮	⋮	⋮	⋮	⋮

x_i – *варианты дискретного вариационного ряда*

n_i – *частоты вариант* x_i

$\bar{x}$ – *выборочная средняя*

S – *выборочное среднее квадратичное отклонение*

h – *шаг*

$\varphi(u_i)$ – *значения табличной функции*

y_i – *выровненные частоты (ординаты) теоретической кривой*

n_i' – *округлённые частоты* y_i *до ближайшего целого числа*

На основании построенных графиков сделать выводы о распределении выборки.

2. Проверить согласованность эмпирического распределения с теоретическим нормальным, применяя четыре критерия:

а) критерий Пирсона;

б) критерий Романовского

б) один из критериев: Колмогорова(чётный вариант), Ястремского; (нечётный)

в) приближенный критерий.

Статистические критерии

Критерий согласия Пирсона χ^2

Критерий согласия Пирсона χ^2 (хи-квадрат) применяют для проверки гипотезы о соответствии эмпирического распределения предполагаемому теоретическому распределению при большом объеме выборки (n > 100) и больших частотах (n_i > 5) вариант x_i

За меру расхождения эмпирического и теоретического распределений английский математик Пирсон принял величину χ^2:

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

где n_i — эмпирические частоты, n_i' — теоретические частоты.

Применение критерия χ^2 к проверке гипотезы о нормальном распределении генеральной совокупности значений признака X осуществляется по следующему правилу.

Правило применения критерия χ^2

1. По имеющейся выборке сделать предположение о нормальном законе распределения признака X генеральной совокупности. Затем найти

оценки параметров этого закона, т.е. найти $\bar{x}, S^2$.

2. Вычислить теоретические частоты n_i по формуле

$$n_i' = \frac{nh}{S}\varphi(u_i)$$

где n — объем выборки, h — шаг, S – выборочное среднее квадратическое

отклонение, $u_i = \frac{x_i - \bar{x}}{S}$, $\varphi(u_i) = \frac{1}{\sqrt{2\pi}} e^{-2u^2/2}$ - находится по таблице, приложение 1.

Для вычисления теоретических частот составить табл.

x_i	$x_i - \bar{x}$	$u_i = \frac{x_i - \bar{x}}{S}$	$\varphi(u_i)$	$n_i' = \frac{nh}{S}\varphi(u_i)$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$

Полученные частоты n_i округлить до целых.

3. Вычислить величину χ^2 по формуле $\chi^2 = \sum_{i=1}^{l} \frac{\left(n_i - n_i'\right)^2}{n_i'}$

и обозначить ее через χ_0^2

Расчет вести, пользуясь.

n_i	n_i'	$n_i - n_i'$	$(n_i - n_i')^2$	$\frac{(n_i - n_i')^2}{n_i'}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
				χ_0^2

4. Найти число степеней свободы ***k*** (параметр распределения Пирсона) по формуле ***k = s - r = s - 3,*** где ***s*** — число интервалов вариационного ряда, ***r*** — сумма числа параметров теоретического закона распределения. Для нормального распределения признака X принято ***r = 3*** (учитываются параметры нормального распределения M_x, σ а также объем выборки ***n***).

5. Выбрать уровень значимости α

6. По найденному числу степеней свободы ***k*** и уровню значимости α, пользуясь таблицей значений Приложение 2 , определить критическое $\chi_{кр}^2$ значение .

Если $\chi_0^2 < \chi_{кр}^2$, то нет достаточных оснований отвергнуть выдвинутую гипотезу о нормальном распределении признака ***X***. Если гипотеза $\chi_0^2 > \chi_{кр}^2$ о нормальном распределении признака ***X*** отвергается.

Критерий Романовского

Для оценки близости эмпирического распределения признака ***X*** к нормальному теоретическому Романовский предложил вычислять отношение:

$$\left|\frac{\chi^2 - k}{\sqrt{2k}}\right|$$

где χ^2 — статистика критерия Пирсона, используя опытные данные, ***k = s – 3*** — число степеней свободы. Если указанное отношение по модулю меньше трех, то расхождение между теоретическим и эмпирическим распределениями считается несущественным, т.е. можно принять, что данное эмпирическое распределение моделируется нормальным распределением.

Если отношение больше трех, то у нас нет оснований считать, что эмпирическое распределение признака ***X*** подчиняется нормальному закону распределения.

Критерий Колмогорова

Критерий Колмогорова в своем классическом виде является более мощным, чем критерий Пирсона, и может быть использован для проверки гипотезы о соответствии эмпирического распределения любому теоретическому непрерывному распределению ***F(x)*** с заранее известными параметрами. Однако параметры функции распределения ***F(x)***, как правило, нам неизвестны, и их оценка производится по данным самой выборки. Это обстоятельство накладывает ограничения на возможность широкого практического применения критерия: он может быть использован только для проверки соответствия опытных данных лишь некоторым конкретным функциям распределения. Для проверки соответствия эмпирического распределения теоретическому нормальному распределению критерий Колмогорова применяют следующим образом.

Вычисляется статистика λ критерия Колмогорова по формуле:

$$\lambda = D / \sqrt{n}$$

Где $D = \max\left|M - M'\right|$ — максимум абсолютного значения разности между накопленными эмпирическими частотами ***M*** и накопленными теоретическими частотами M' ***n***— объем выборки.

По вычисленному λ находят значение функции:

$$K(\lambda) = 1 - \sum_{-\infty}^{\infty} (-1)^k e^{-2k^2\lambda^2}$$

— вероятность того, что λ достигает данной величины.

Если найденному значению λ соответствует очень малая вероятность, т.е. $K(\lambda) < 0{,}05$, то имеет место существенное расхождение между эмпирическим и теоретическим распределениями, которое нельзя считать случайным. Следовательно, рассматриваемая выборка не может быть смоделирована нормальным законом распределения. Если вероятность $K(\lambda) > 0{,}05$, то расхождение между частотами может быть случайным, и распределения хорошо соответствуют одно другому.

Значения $K(\lambda)$ находят по таблице

λ	$K(\lambda)$	λ	$K(\lambda)$
0,30	1,0000	1,10	0,1777
0,35	0,9997	1,20	0,1122
0,40	0,9972	1,30	0,681
0,45	0,9874	1,40	0,397
0,50	0,9639	1,50	0,222
0,55	0,9228	1,60	0,120
0,60	0,8643	1,70	0,052
0,70	0,7112	1,90	0,015
0,75	0,6272	2,00	0,007
0,80	0,5441	2,10	0,0003
0,85	0,4653	2,20	0,0001
0,90	0,3927	2,30	0,0001
0,95	0,3275	2,40	0,0000
1,00	0,2700	2,50	0,0000

Критерий Ястремского

Для проверки соответствия данной выборочной совокупности признака ***X*** нормальному распределению составляется неравенство:

$$J \le 3\sqrt{2l + 4Q},$$

$$где\, J = |c - l|;$$

$$c = \sum \frac{(n_i - n_i')^2}{n_i' \, q_i'}$$

$$n_i - эмпирические \quad частоты$$

$$n_i' - теоретические\, частоты$$

$$l - число\, столбцов\, дискретного\, вариационного\, ряда$$

$$q_i' = 1 - p_i';\, p' = \frac{n_i'}{n};\, n - обьём\ выборки.\ Если\ l < 20, то\ Q = 0.6$$

Если вычисленное значение ***J*** меньше $3\sqrt{2l + 4Q},$ то гипотеза о близости эмпирического распределения признака ***X*** к нормальному закону распределения принимается. При значениях ***J*** больших $3\sqrt{2l + 4Q},$

расхождения между эмпирическим и теоретическим распределениями существенно. В этом случае данные выборки не будут подчиняться нормальном закону распределения.

Для вычисления величины ***c*** составляется табл.

x_i	n_i	n'_i	p'_i	q'_i	$n_i - n'_i$	$(n_i - n'_i)^2$	$n'_i q'_i$	$\dfrac{(n_i - n'_i)^2}{n'_i q'_i}$
$\sum$								$c =$

Приближенные критерии нормальности распределения

Для проверки гипотезы о соответствии данной выборки нормальному закону распределения используют выборочные статистики: асимметрию и эксцесс. Их вычисляют по формулам из Лаб.1.

Затем вычисляют их средние квадратические отклонения по формулам:

$$S_{A_s} = \sqrt{\frac{6(n-1)}{(n+1)(n+3)}};$$

$$S_{E_x} = \sqrt{\frac{24n(n-2)(n-3)}{(n-1)^2(n+3)(n+5)}}.$$

Если $|A_s| < S_{A_s}, |E_x| < S_{E_x}$, то выборочная совокупность подчиняется нормальному закону распределения. Если и заметно больше своих средних квадратических отклонений, то выборочная совокупность не будет распределена по нормальному закону.

Контрольные вопросы.

1. Рассказать о возможных вариантах построения кривой нормального распределения по опытным данным.

2. Дать определение статистической гипотезы.

3. Что называется статистическим критерием?

4. Описать алгоритм применения любого статистического критерия для обработки экспериментальных данных.

5. Сформулировать правило применения критерия согласия Пирсона для проверки гипотезы согласованности эмпирического распределения с теоретическим нормальным.

6. Рассказать о применении критерия согласия Романовского для оценки близости эмпирического распределения к теоретическому нормальному.

7. Описать алгоритм применения критерия Колмогорова для проверки соответствия эмпирического распределения нормальному теоретическому распределению.

8. Рассказать о применении критерия Ястремского для проверки соответствия данной выборочной совокупности нормальному распределению.

9. Рассказать о приближенных критериях, применяемых для проверки гипотезы о нормальном распределении выборочной совокупности.

Приложение 1

Таблица значений функции $\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$

x	0	1	2	3	4	5	6	7	8	9
0,0	0,3989	3989	3989	3988	39986	3984	3982	3980	3977	3973
0,1	0,3970	3965	3961	3956	3951	3945	3939	3932	3925	3918
0,2	0,3910	3902	3894	3885	3876	3867	3857	3847	3836	3825
0,3	0,3814	3802	3790	3778	3765	3752	3739	3726	3712	3697
0,4	0,3683	3668	3653	3637	3621	3605	3589	3572	3555	3538
0,5	0,3521	3503	3485	3467	3448	3429	3410	3391	3372	3352
0,6	0,3332	3312	3292	3271	3251	3230	3209	3187	3166	3144
0,7	0,3123	3101	3079	3056	3034	3011	2989	2966	2943	2920
0,8	0,2897	2874	2850	2827	2803	2780	2756	2732	2709	2685
0,9	0,2661	2637	2613	2589	2565	2541	2516	2492	2468	2444
1,0	0,2420	2396	2371	2347	2323	2299	2275	2251	2227	2203
1,1	0,2179	2155	2131	2107	2083	2059	2036	2012	1989	1965
1,2	0,1942	1919	1895	1872	1849	1826	1804	1781	1758	1736
1,3	0,1714	1691	1669	1647	1626	1604	1582	1561	1539	1518
1,4	0,1497	1476	1456	1435	1415	1394	1374	1354	1334	1315
1,5	0,1295	1276	1257	1238	1219	1200	1182	1163	1145	1127
1,6	0,1109	1092	1074	1057	1040	1023	1006	0989	0973	0957
1,7	0,0940	0925	0909	0893	0878	0863	0848	0833	0818	0804
1,8	0,0790	0775	0761	0748	0734	0721	0707	0694	0681	0669
1,9	0,0656	0644	0632	0620	0608	0596	0584	0573	0562	0551
2,0	0,0540	0529	0519	0508	0498	0488	0478	0468	0459	0449
2,1	0,0440	0431	0422	0413	0404	0396	0387	0379	0371	0363
2,2	0,0355	0347	0339	0332	0325	0317	0310	0303	0297	0290
2,3	0,0283	0277	0270	0264	0258	0252	0246	0241	0235	0229
2,4	0,0224	0219	0213	0208	0203	0198	0194	0189	0184	0180
2,5	0,0175	0171	0167	0163	0158	0154	0151	0147	0143	0139
2,6	0,0136	0132	0129	0126	0122	0119	0116	0113	0110	0107
2,7	0,0104	0101	0099	0096	0093	0091	0088	0086	0084	0081
2,8	0,0079	0077	0075	0073	0071	0069	0067	0065	0063	0061
2,9	0,0060	0058	0056	0055	0053	0051	0050	0048	0047	0046
3,0	0,0044	0043	0042	0040	0039	0038	0037	0036	0035	0034
3,1	0,0033	0032	0031	0030	0029	0028	0027	0026	0025	0025
3,2	0,0024	0023	0022	0022	0021	0020	0020	0019	0018	0018
3,3	0,0017	0017	0016	0016	0015	0015	0014	0014	0013	0013
3,4	0,0012	0012	0012	0011	0011	0010	0010	0010	0009	0009
3,5	0,0009	0008	0008	0008	0008	0007	0007	0007	0007	0006
3,6	0,0006	0006	0006	0005	0005	0005	0005	0005	0005	0004
3,7	0,0004	0004	0004	0004	0004	0004	0003	0003	0003	0003
3,8	0,0003	0003	0003	0003	0003	0002	0002	0002	0002	0002
3,9	0,0002	0002	0002	0002	0002	0002	0002	0002	0001	0001

Приложение 2.

Критические точки распределения χ^2

Число степеней свободы k	Уровень значимости α					
	0,01	0,025	0,05	0,95	0,975	0,98
1	6,6	5,024	3,841	0,0039	0,00098	0,00016
2	9,2	7,378	5,991	0,103	0,051	0,020
3	11,3	9,348	7,815	0,352	0,216	0,115
4	13,3	11,143	9,488	0,711	0,484	0,297
5	15,1	12,832	11,070	1,15	0,831	0,554
6	16,8	14,449	12,592	1,64	1,24	0,872
7	18,5	16,013	14,067	2,17	1,69	1,24
8	20,1	17,535	15,507	2,73	2,18	1,65
9	21,7	19,023	16,919	3,33	2,70	2,09
10	23,2	20,483	18,307	3,94	3,25	2,56
11	24,7	21,920	19,676	4,57	3,82	3,05
12	26,2	23,336	21,026	5,23	4,40	3,57
13	27,7	24,736	22,362	5,89	5,01	4,11
14	29,1	26,129	23,685	6,57	5,63	4,66
15	30,6	27,488	24,996	7,26	6,26	5,23
16	32,0	28,845	26,296	7,96	6,91	5,81
17	33,4	30,191	27,587	8,67	7,56	6,41
18	34,8	31,536	28,869	9,39	8,23	7,01
19	36,2	32,852	30,144	10,1	8,91	7,63
20	37,6	34,170	31,410	10,9	9,59	8,26
21	38,9	35,479	32,671	11,6	10,3	8,90
22	40,3	36,781	33,924	12,3	11,0	9,54
23	41,6	38,076	35,172	13,1	11,7	10,2
24	43,0	39,364	36,415	13,8	12,4	10,9
25	44,3	40,646	37,652	14,6	13,1	11,5
26	45,6	41,923	38,885	15,4	13,8	12,2
27	47,0	43,194	40,113	16,2	14,6	12,9
28	48,3	44,461	41,337	16,9	15,3	13,6
29	49,6	45,722	42,557	17,7	16,0	14,3
30	50,9	46,979	43,773	18,5	16,8	15,0