ДОНСКОЙ ГОСУДАРСТВЕННЫЙ ТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ

УПРАВЛЕНИЕ ДИСТАНЦИОННОГО ОБУЧЕНИЯ И ПОВЫШЕНИЯ КВАЛИФИКАЦИИ

Кафедра «Технология вяжущих веществ, бетонов и строительной керамики»

Методические указания

для выполнения практической работы

«Статистическая обработка результатов испытаний. Проверка нормальности распределения»

Автор
Серебряная И.А.

Ростов-на-Дону, 2017

Аннотация

Содержатся сведения о возможности обработки экспериментальных данных, а также проверки гипотезы о нормальности распределения измеряемой величины с помощью статистических методов. Предназначены для обучающихся по направлению подготовки 27.03.01 «Стандартизация и метрология», 27.04.02 «Управление качеством», 08.04.01 «Строительство».

Автор

к.т.н., доцент
кафедры «ТВВБиСК»
Серебряная И.А.

Оглавление

1. ОБЩАЯ ЧАСТЬ

1.1. В теории вероятностей и математической статистике часто целесообразно связывать различные события с действительными числами и вести расчеты последних. В результате этого следует ввести понятие: *случайная величина (случайная переменная)* – это величина, которая при любом испытании случайным образом принимает любое действительное числовое значение.

Дискретная случайная величина X – величина, которая может принимать конечное или счетное число различных значений $x_1, x_2, \ldots, x_n$.

НАПРИМЕР:

– при подсчете простоев станка за смену X может принимать значения 0, 1, 2, … ;

– при подсчете числа образцов X может принимать значения 0, 1, 2, … .

Непрерывная случайная величина X – величина, которая обладает следующим свойством, а именно, она может принимать любые значения в определенном интервале числовой оси.

НАПРИМЕР: – при замере прочности бетонных образцов X может принимать любое значение в интервале, ограниченном заданным допуском, например от 32 до 35 МПа. Причем данное высказывание ограничивается несовершенством измерительных средств, допускающих лишь определенную точность измерений.

1.2. Для полного определения *случайной величины X* важно наряду со знанием области изменения ее значений знать, *как часто* эти значения встречаются в большой серии опытов. *Закон распределения случайной величины* – всякое соотношение, устанавливающее связь между возможны значениями случайной величины и соответствующими им вероятностями.

Так для *дискретной случайной величины* – это вероятность (P_i) всех значений, которые может принимать X_i (ряд распределения). Для *непрерывной случайной величины* невозможно полностью указать все вероятности, необходимые для характеристики случайной величины X, т.к. не всякое бесконечное множество можно представить в виде последовательности. В то же время различные области возможных значений случайной непрерывной величины не являются одинаково вероятными, поэтому и для непрерывной величины существует распределение вероятностей. Вероятность непрерывного события X есть функция от X. Эта функция называется *функцией распределения (F(x))*.

1.3. *Функция распределения случайной величины X* определяется выражением: *F(x)= P(X<x),* где x пробегает все без исключения значения на действительной числовой оси. Это выражение имеет следующий смысл: функция распределения случайной величины X равна вероятности того, что X принимает значения ниже предела x.

Функцию распределения называют так же и интегральной функцией или интегральным законом распределения величины:

$$F(x) = \int_{-\infty}^{x} \varphi(x)dx, \qquad (1)$$

где φ(x) (плотность распределения) – производная F(x) и характеризует плотность, с которой распределяются значения случайной непрерывной величины.

Функция распределения представляет собой некоторую абстрактную математическую модель, при помощи которой описываются случайные экспериментально наблюдаемые величины.

Одна из задач статистической обработки материала (случайной величины X) заключается в нахождении такой функции распределения, которая описывала бы достаточно хорошо наблюдаемые значения случайной величины и была бы удобна для статистического анализа.

1.4. Наиболее известным и часто применяемым при решении задач математической статистики и статистического контроля качества, особенно когда имеет место большое количество измерений ($n > 50$) распределением случайной величины является *нормальное распределение.*

Нормальный закон распределения именуют законом Гаусса. Главная особенность нормального закона состоит в том, что он является предельным законом, к которому приближаются другие законы распределения (экспоненциальное, Пуассона, геометрическое, биномиальное и т.д.).

Нормальный закон распределения характеризуется *плотностью вероятности*:

$$\varphi(x) = \frac{1}{\sigma_x\sqrt{2\pi}} exp^{-\frac{(x_i - x_0)^2}{2\sigma_x^2}}, \qquad (2)$$

где exp=2,718…, п = 3,142…, σ и x_0 – параметры нормального распределения (соответственно, среднеквадратической отклонение и математическое ожидание).

2. ПРОВЕРКА НОРМАЛЬНОСТИ РАСПРЕДЕЛЕНИЯ РЕЗУЛЬТАТОВ НАБЛЮДЕНИЙ

Итак, одной из важнейших задач анализа вариационных рядов является выявление закономерности распределения и определение ее характера. Основной путь выявления закономерности распределения – построение вариационных рядов.

Отклонение от нормального распределения может быть случайным или неслучайным. Если это отклонение неслучайно, то возникает необходимость изменять методику определения измеряемой величины. Если отклонение случайное – необходимо определить степень отклонения от нормального распределения или выяснить, какому иному (t-распределению, F-распределению) оно подчиняется.

В статистике разработан ряд критериев для оценки степени близости наблюдаемого распределения к нормальному.

2.1. Графический способ определения нормальности распределения

Наиболее простым, но весьма приближенным методом оценки согласия результатов эксперимента с тем или иным законом распределения является графический метод. Опытные данные наносят на вероятностную бумагу и сравнивают с графиком принятой функции распределения, которая на вероятностной сетке изображается прямой линией. Если экспериментальные точки ложатся вблизи прямой со случайными отклонениями влево и вправо, то опытные данные соответствуют рассматриваемому закону распределения (рис. 1 и рис. 2). Систематическое и значительное отклонения экспериментальных точек от аппроксимирующей прямой свидетельствует об ошибочности принятой модели для обоснования закона распределения исследуемой случайной величины (рис.3).

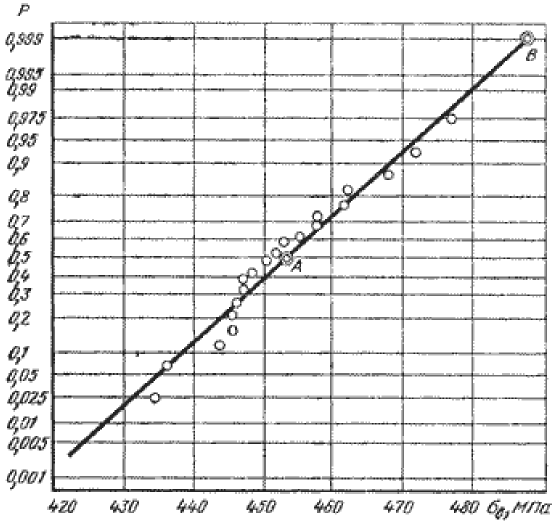

Рис. 1 График эмпирической функции распределения предела прочности образцов из дюралюминиевого прессованного профиля на нормальной вероятностной сетке

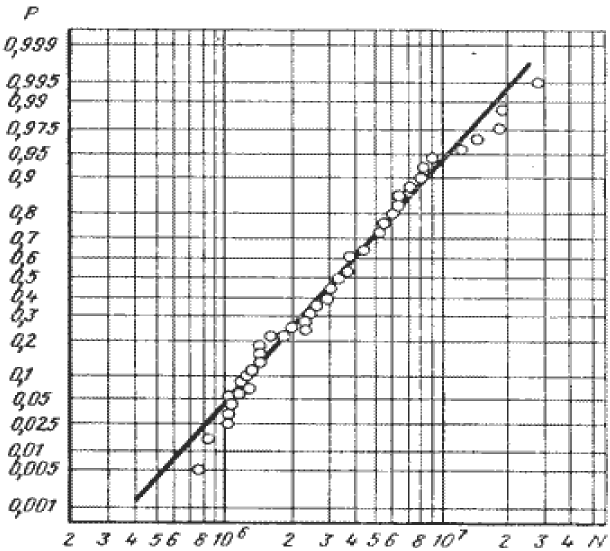

Рис. 2 График эмпирической функции распределения числа циклов до разрушения образцов на логарифмической нормальной вероятностной сетке

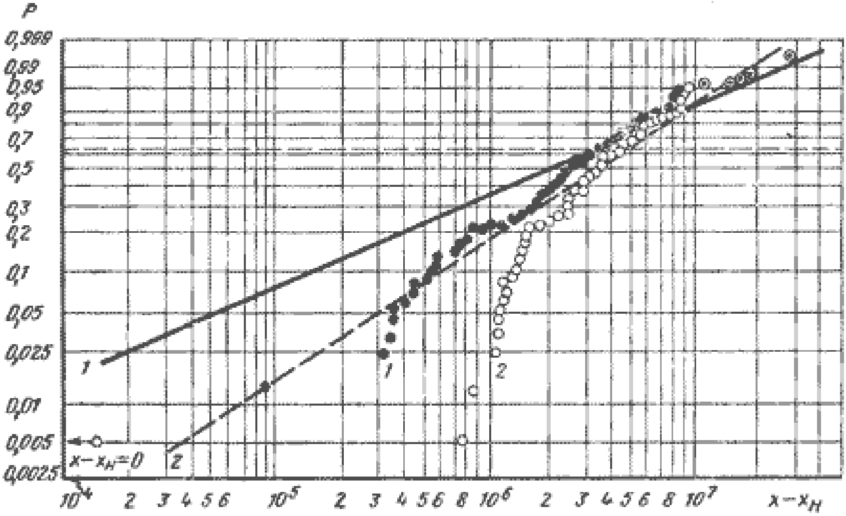

Рис. 3 График эмпирической функции распределения числа циклов до разрушения образцов на вероятностной сетке распределения Вейбула – Гнеденко: 1 – трехпараметрического, 2 – двухпараметрического.

Графический способ в значительной степени является субъективным и используется на практике лишь в качестве первого приближения при решении этой задачи.

2.2. Графический способ определения нормальности распределения (гистограмма).

2.2.1 При большом числе результатов наблюдений (n>40) данная задача решается в следующем порядке.

– Весь диапазон полученных результатов наблюдений $X_{min}...X_{max}$ разделяют на r интервалов шириной d_x (i=1,2,…,r) и подсчитывают частоты m_i, равные числу результатов, лежащих в каждом i-м интервале, т. е. меньших или равных его правой и больших левой границы.

$$d_x = \frac{X_{max} - X_{min}}{2}, \qquad (3)$$

– Отложим вдоль оси результатов наблюдений интервалы d_x в порядке возрастания индекса i и на каждом интервале построим прямоугольник с высотой, равной m_i. Полученный график называется гистограммой статистического распределения.

При увеличении числа наблюдений число интервалов можно увеличить. Сами интервалы уменьшаются, и гистограмма все больше приближается к плавной кривой, ограничивающей единичную площадь, – к графику плотности распределения результатов наблюдений.

2.2.2 При построении гистограмм рекомендуется пользоваться *следующими правилами:*

– Число интервалов выбирается в зависимости от числа наблюдений согласно рекомендациям таблицы 1.

Таблица 1

количество измерений (n)	число интервалов (r)
40 – 100	7 – 9
100 – 500	8 – 12
500 – 1000	10 – 16
1000–10000	12 – 22

– Длины интервалов удобнее выбирать одинаковыми. Однако если распределение крайне неравномерно, то в области максимальной концентрации результатов наблюдений следует выбирать более узкие интервалы.

– Масштабы по осям гистограммы должны быть такими, чтобы отношение ее высоты к основанию составляло примерно 5:8.

2.2.3 После построения гистограммы надо подобрать теоретическую плавную кривую распределения, которая, выражая все существенные черты статистического распределения, сглаживала бы все случайности, связанные с недостаточным объемом экспериментальных данных. Принципиальный вид теоретической кривой выбирают заранее, проанализировав метод измерения, или хотя бы по внешнему виду гистограммы. Тогда определение аналитического вида кривой распределения сводится к выбору таких значений его параметров, при которых достигается наибольшее соответствие между теоретическим и статистическим распределением.

2.2.4 *Задача сглаживания* сводится к вычислению теоретических частот распределения и построению по ним выравнивающей кривой распределения.

Теоретические частоты рассчитываются по формуле (с округлением до целых):

$$m_i^T = \frac{nd_x}{S_B}\varphi(u_i), \qquad (4)$$

где u_i - аргумент дифференциальной функции нормированного распределения.

$$u_i = \frac{x_к - \overline{X}_B}{S_B}, \qquad (5)$$

где $x_к$ – среднее значение в интервале.

Значения $\varphi(u_i)$ принимаются по справочной таблице (Приложение 1).

Расчет теоретических частот целесообразно выполнять в табличной форме (таблица 2).

Таблица 2

Интервал	Среднее значение в интервале, $X_к$	Частота, m_i	U_i	$\varphi(u_i)$	Теоретическая частота, m_i^T
1					
2					
…					
к					

Для наглядности полученных данных строят статистические распределения экспериментальных и теоретических частот. Для этого на график (гистограмму) по оси абсцисс откладываем средние значения интервала ($X_к$), а по оси ординат - частоты (m_i, m_i^T).

При визуальном сравнении если графики приблизительно совпадают, то можно считать, что закон близок, к нормальному.

2.3. Проверка гипотезы о нормальности распределения.

Далее законно возникает вопрос, объясняются ли расхождения между гистограммой (m_i) и подобранным теоретическим распределением (m_i^T) только случайными обстоятельствами, связанными с ограниченным числом наблюдений, или они вызваны тем, что результаты наблюдений в действительности распределены иначе?

Для ответа на этот вопрос используют методы проверки статистических гипотез. Идея их применения заключается в следующем.

2.3.1 На основании гистограммы, полученной при обработке опытных данных, строится гипотеза, состоящая в том, что результаты наблюдений подчиняются распределению $F_X(x)$ с плотностью $P_X(x)$.

Если закон распределения неизвестен, но есть основания предположить, что он имеет определенный вид (назовем его А), то проверяют нулевую гипотезу H_0: генеральная совокупность распределена по закону А. H_1: генеральная совокупность не распределена по закону А.

Проверку гипотезы о предполагаемом законе неизвестного распределения производится так же, как и проверка гипотезы о параметрах распределения, т.е. при помощи специально подобранной случайной величины – критерия согласия.

Критерием согласия называют критерий проверки гипотез о предполагаемом законе неизвестного распределения.

Имеется несколько критериев согласия Пирсона ***(χ^2)***, Колмогорова-Смирнова и др.

2.3.2 Критерий Пирсона χ^2

*Критерий согласия Пирсона (**χ^2**)* применяют для проверки гипотезы о соответствии эмпирического распределения предполагаемому теоретическому распределению F(x) при большом объеме выборки ($n \geq 100$). Критерий применим для любых видов функции F(x), даже при неизвестных значениях их параметров, что обычно имеет место при анализе результатов механических испытаний. В этом заключается его универсальность.

В этом разделе студенты самостоятельно, используя учебник В.Е. Гмурмана "Теория вероятностей и математическая статистика", проводят оценку нормальности распределения с *помощью критерия Пирсона. Предварительно рекомендуется рассчитать теоретические частоты* (m_i^T)*в соответствии с п. 2.2.4.*

2.3.3 Критерий Колмогорова – Смирнова λ .

Этот критерий можно применять при малых выборках. Данный метод уступает по точности методу χ^2 -Пирсона, но гораздо проще в применении.

Данные должны представлять случайную выборку, переменные должны быть измерены по порядковой шкале, должна быть сформулирована гипотеза о распределении генеральной совокупности.

В основе оценки распределения с помощью λ-критерия лежит теорема А.Н. Колмогорова о распределении максимума отклонений теоретической интегральной функции распределения от соответствующей эмпирической функции.

Критерий λ определяется по выражению:

$$D_{max} = \frac{\max|N - N'|}{n}, \qquad \lambda = D_{max}\sqrt{n}, \qquad (6)$$

где D_{max} – максимальное отклонение теоретической интегральной функции распределения от соответствующей эмпирической функции;

N – накопленные эмпирические частоты;

N′ - накопленные теоретические частоты;

n – число измерений.

Далее по Приложению 2, исходя из рассчитанного λ, определяют вероятность P(λ). Исходя из практического опыта можно считать расхождение между эмпирическим и теоретическим нормальным распределением незначительным, уже при $P(\lambda) \geq 0,6$. При значении $P(\lambda) < 0,05$ – расхождение между эмпирическим и теоретическим нормальным распределением признают неслучайным, а значит, распределение не подчиняется нормальному закону. Если $0,05 < P(\lambda) < 0,6$, значит в данном случае критерий λ не подходит для оценки нормальности распределения.

ЛИТЕРАТУРА

1. Гмурман В.Е. Теория вероятностей и математическая статистика: Учебник. – М.: Высшая школа, 2004 – 407 с.
2. Елисеева И.И., Юзбашев М.М. Общая теория статистики: Учебное пособие. – М.: Финансы и статистика, 2004. – 656 с.
3. Зазимко В.Г. Оптимизация свойств строительных материалов: Учебное пособие. – М.: Транспорт, 1981. – 103 с.
4. Шторм Р. Теория вероятностей, математическая статистика, статистический контроль качества. –М.: Мир, 1970. – 368 с.

ПРИЛОЖЕНИЕ 1

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

	0	1	2	3	4	5	6	7	8	9
0,0	0,3989	3989	3989	3988	3986	3984	3982	3980	3977	3973
0,1	3970	3965	3961	3956	3951	3945	3939	3932	3925	3918
0,2	3910	3902	3894	3885	3876	3867	3857	3847	3836	3825
0,3	3814	3802	3790	3778	3765	3752	3739	3726	3712	3697
0,4	3683	3668	3652	3637	3621	3605	3589	3572	3555	3538
0,5	3521	3503	3485	3467	3448	3429	3410	3391	3372	3352
0,6	3332	3312	3292	3271	3251	3230	3209	3187	3166	3144
0,7	3123	3101	3079	3056	3034	3011	2989	2966	2943	2920
0,8	2897	2874	2850	2827	2803	2780	2756	2732	2709	2685
0,9	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	2179	2155	2131	2107	2083	2059	2036	2012	1989	1965
1,2	1942	1919	1895	1872	1849	1826	1804	1781	1758	1736
1,3	1714	1691	1669	1647	1626	1604	1582	1561	1539	1518
1,4	1497	1476	1456	1435	1415	1394	1374	1354	1334	1315
1,5	1295	1276	1257	1238	1219	1200	1182	1163	1145	1127
1,6	1109	1092	1074	1057	1040	1023	1006	0989	0973	0957
1,7	0940	0925	0909	0893	0878	0836	0848	0833	0818	0804
1,8	0790	0775	0761	0748	0734	0721	0707	0694	0681	0669
1,9	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	0440	0431	0422	0413	0404	0396	0387	0379	0371	0363
2,2	0355	0347	0339	0332	0325	0317	0310	0303	0297	0290
2,3	0283	0277	0270	0264	0258	0252	0246	0241	0235	0229
2,4	0224	0219	0213	0208	0203	0198	0194	0189	0284	0180
2,5	0175	0171	0267	0163	0158	0154	0151	0147	0143	0139
2,6	0136	0132	0129	0126	0122	0119	0116	0113	0110	0108
2,7	0104	0101	0099	0096	0193	0091	0088	0086	0084	0081
2,8	0079	0077	0075	0073	0171	0069	0067	0065	0063	0061
2,9	0060	0058	0056	0055	0153	0051	0050	0048	0047	0046
3,0	0,0044	0043	0042	0040	0039	0038	0037	0036	0035	0034
3,1	0033	0032	0031	0030	0029	0028	0027	0026	0025	0025
3,2	0024	0023	0022	0022	0021	0020	0020	0019	0018	0018
3,3	0017	0017	0016	0016	0015	0015	0041	0014	0013	0013
3,4	0012	0012	0012	0011	0011	0010	0010	0010	0009	0009
3,5	0009	0008	0008	0008	0008	0007	0007	0007	0007	0006
3,6	0006	0006	0006	0005	0005	0005	0005	0005	0005	0004
3,7	0004	0004	0004	0004	0004	0004	0003	0003	0003	0003
3,8	0003	0003	0003	0003	0003	0002	0002	0002	0002	0002
3,9	0002	0002	0002	0002	0002	0002	0002	0002	0001	0001

ПРИЛОЖЕНИЕ 2

Вероятность P(λ) для различных значений λ
(распределение А.Н.Колмогорова)

λ	P(λ)	λ	P(λ)	λ	P(λ)
0,30	1,0000	0,80	0,5441	1,60	0,0120
0,35	0,9997	0,85	0,4653	1,70	0,0062
0,40	0,9972	0,90	0,3927	1,80	0,0032
0,45	0,9874	0,95	0,3275	1,90	0,0015
0,50	0,9639	1,00	0,2700	2,00	0,0007
0,55	0,9228	1,10	0,1777	2,10	0,0003
0,60	0,8643	1,20	0,1122	2,20	0,0001
0,65	0,7920	1,30	0,0681	2,30	0,0001
0,70	0,7112	1,40	0,0397	2,40	0,0000
0,75	0,6272	1,50	0,0222	2,50	0,0000