Лабораторная работа № 8

ПРОВЕРКА ГИПОТЕЗЫ О СООТВЕТСТВИИ ВЫБОРКИ НОРМАЛЬНОМУ ЗАКОНУ РАСПРЕДЕЛЕНИЯ

Цель работы: приобрести навыки применения статистических гипотез о виде неизвестного закона распределения по критерию Пирсона.

Основные теоретические сведения

С теорией статистического оценивания параметров тесно связана проверка статистических гипотез. Она используется всякий раз, когда необходим обоснованный вывод о преимуществах того или иного способа инвестиций, измерений, технологического процесса, об эффективности нового метода обучения, управления, об уровне доходности ценных бумаг, о значимости математической модели и т.д.

Статистической гипотезой называется любое предположение о виде или параметрах неизвестного закона распределения. Проверяемую гипотезу обычно называют нулевой (или основной) и обозначают H_0. Наряду с нулевой гипотезой H_0 рассматривают альтернативную, или конкурирующую, гипотезу H_1, являющуюся логическим отрицанием H_0. Нулевая и альтернативная гипотезы представляют собой две возможности выбора, осуществляемого в задачах проверки статистических гипотез.

Правило, по которому гипотеза H_0 отвергается или принимается, называется статистическим критерием или статистическим тестом. При этом возможны четыре случая, отображенные в табл. 8.1.

Таблица 8.1

Гипотеза H_0	Принимается	Отвергается
Верна	Правильное решение	Ошибка 1-го рода
Неверна	Ошибка 2-го рода	Правильное решение

Вероятность α допустить ошибку 1-го рода, т.е. отвергнуть гипотезу H_0, когда она верна, называется уровнем значимости (или размером) критерия.

Вероятность допустить ошибку 2-го рода, т.е. принять гипотезу H_0, когда она неверна, обычно обозначают β.

Вероятность (1 – β) не допустить ошибку 2-го рода, т.е. отвергнуть гипотезу H_0, когда она неверна, называется мощностью критерия.

Пользуясь терминологией статистического контроля качества продукции, можно сказать, что вероятность α представляет «риск поставщика», связанный с забраковкой по результатам выборочного контроля изделий всей партии, удовлетворяющей стандарту, а вероятность β – «риск потребителя», связанный с принятием по анализу выборки партии, не удовлетворяющей стандарту.

Методы проверки статистических гипотез занимают центральное место в исследованиях математической статистики. Одной из важнейших групп критериев проверки статистических гипотез являются критерии их проверки о виде распределений (критерии согласия). Они по выборочным данным проверяют предположение о принадлежности генеральной совокупности к тому или иному виду распределений.

Одним из наиболее мощных критериев согласия является критерий Пирсона, называемый еще критерием хи-квадрат (χ^2). По результатам экспериментальных измерений и предположению нормального закона их распределения определяется расчетное значение критерия Пирсона.

$$\chi^2 = \sum_{i=1}^{k} \frac{\left(n_i - np_i\right)^2}{np_i}.$$

Далее вычисляют число степеней свободы, задаются уровнем значимости α и определяют критическое значение критерия Пирсона $\chi^2_{\alpha,\,\nu}$ со степенями свободы $\nu = (k - r - 1)$, где r – число неизвестных параметров закона распределения, оцениваемых по выборке, k – число интервалов группированного статистического ряда.

Если $\chi^2 < \chi^2_{\alpha,\,\nu}$, то нуль-гипотеза о нормальном законе распределения экспериментальных данных принимается с доверительной вероятностью $p = 1 - \alpha$. В противном случае нуль-гипотеза отвергается и принимается альтернативная.

Задания на выполнение лабораторной работы

1. Освойте алгоритм работы с функциями MS Excel.
2. Проверьте согласованность эмпирического распределения с теоретическим нормальным по критерию Пирсона.
3. Постройте эмпирическую (полигон) и теоретическую (нормальную) кривые распределения.
4. Ответьте на контрольные вопросы.

Методика выполнения работы

1. Перед выполнением данной работы ознакомьтесь с функциями MS Excel, которые применяются для определения критических значений критерия Пирсона (хи-квадрат) и функции Лапласа.

Встроенная функция ХИ2ОБР используется для определения теоретического значения критерия Пирсона (рис. 8.1).

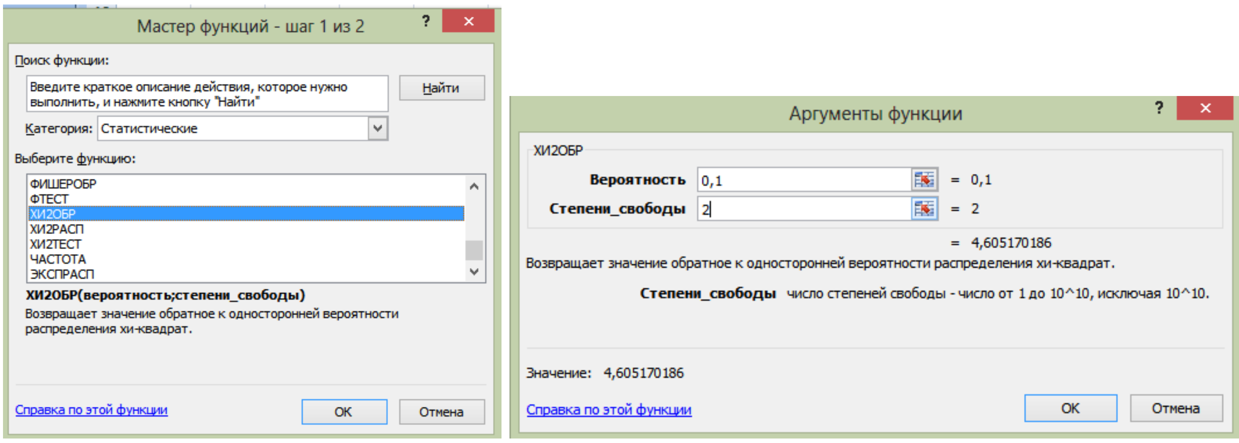


Рис. 8.1. Встроенная функция ХИ2ОБР

Для определения функции Лапласа применяется встроенная функция НОРМСТРАСП.

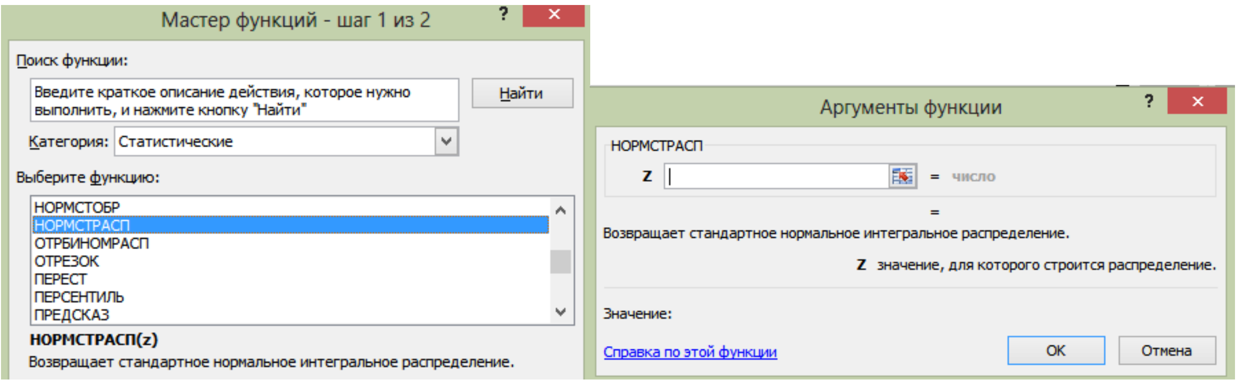
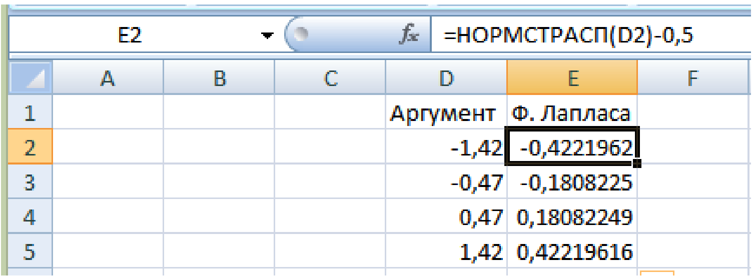


Рис. 8.2. Встроенная функция НОРМСТРАСП

Название функции НОРМСТРАСП является сокращением от термина «нормальное стандартное распределение», так как ее главной задачей является возврат в выделенную ячейку стандартного нормального интегрального распределения. Синтаксис функции:

= НОРМСТРАСП (x; интегральная (1)).

Для того чтобы выполнить вычисление функции Лапласа от аргумента *x* применяется следующая формула: = НОРМСТРАСП(x) – 0,5 (рис.8.3).



E2 =НОРМСТРАСП(D2)-0,5

	A	B	C	D	E	F
1				Аргумент	Ф. Лапласа	
2				-1,42	-0,4221962	
3				-0,47	-0,1808225	
4				0,47	0,18082249	
5				1,42	0,42219616	

Рис. 8.3. Пример вычисления функции Лапласа

2. Создайте шаблон таблицы MS Excel и в ячейки A1: A100 введите двухзначные целые числа, созданные через генератор случайных чисел (n = 100).

3. По методике описанной в работе № 7 выполните группировку исходного массива данных и определите частоту попадания элементов массива в соответствующий интервал, а также выборочное среднее ($\bar{x}$) и выборочное среднее квадратическое отклонение (S).

4. Введите в таблицу параметры, которые необходимы для вычисления критерия Пирсона :

– частоту попаданий элементов выборки в i-й интервал, n_i;

– левую границу i-го интервала, x_i ;

– правую границу i-го интервала, x_{i+1};

– нормированные значения левой границы интервала, $z_i = \dfrac{x_i - \bar{x}}{S}$;

– нормированные значения правой границы интервала, $z_{i+1} = \dfrac{x_{i+1} - \bar{x}}{S}$;

– значения функции Лапласа для соответствующих аргументов $\Phi(z)$;

– вероятность попадания СВ в интервал $p_i = \Phi(z_{i+1}) - \Phi(z_i)$;

– теоретическое значение частот $n_i' = np_i$, где n – объем выборки.

Значения функции Лапласа $\Phi(z)$ определите с помощью встроенной функции НОРМСТРАСП (рис. 8.2, 8.3).

Пример представления результатов расчета показан в табл. 8.2

Таблица 8.2

x_i	x_{i+1}	n_i	z_i	z_{i+1}	$\Phi(z_i)$	$\Phi(z_{i+1})$	p_i	np_i	$\dfrac{(n_i - np_i)^2}{nP_i}$
-20	-10	19		-1,42	-0,5	-0,4222	0,0778	15,56	0,761
-10	0	42	-1,4,	-0,47	-0,4222	-0,1808	0,2414	48,28	0,817
0	10	71	-0,47	0,47	-0,1808	0,1808	0,3616	72,32	0,024

10	20	56	0,47	1,42	0,1808	0,4222	0,2414	48,28	1,234
20	30	12	1,42		0,4222	0,5	0,0778	15,56	0,814
Сумма		200,000					1,000	200,000	3,650

Таким образом, получили расчетное значение критерия Пирсона:

$$\chi^2 = \sum_{i=1}^{k} \frac{\left(n_i - np_i\right)^2}{np_i} = 3{,}650.$$

По встроенной функции ХИ2ОБР (рис. 8.1) определите значение критической точки распределения χ^2 при уровне значимости $\alpha = 0{,}05$ и числу степеней свободы $\nu = (5 - 2 - 1) = 2$.

Для данного примера получили, что $\chi^2_{кр} = 6{,}0$. Так как расчетное значение $\chi^2 < \chi^2_{кр}$, то следует принять гипотезу о нормальном распределении исследуемой выборки с доверительной вероятностью $p = 1 - \alpha$.

7. Проверьте гипотезу для уровней значимости $\alpha = 0{,}1$ и $\alpha = 0{,}5$.

8. Постройте совместно эмпирическую и теоретическую кривые распределения и сделайте общий вывод по работе (пример на рис. 8.4).

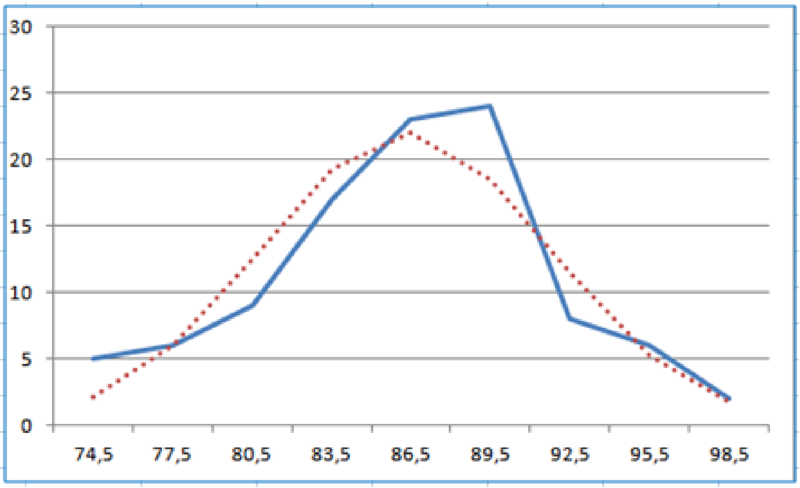


Рис. 8.4. Сравнение эмпирической и теоретической (пунктиром) кривых распределения

Задания для индивидуального выполнения

Создайте шаблон MS Excel и в ячейки А1: А100 введите двухзначные целые числа, созданные через генератор случайных чисел (n = 100).

Затем выполните работу по вышеописанному алгоритму.

Требования к оформлению отчета

Отчет должен содержать следующие составляющие:

1. Тема лабораторной работы.
2. Цель работы.
3. Основные формулы, методика выполнения и алгоритм расчетов.
4. Полученные результаты в виде таблиц, графиков.
5. Анализ полученных результатов.
6. Вывод по работе.

Контрольные вопросы

1. Что называется статистической гипотезой?
2. Что такое альтернативная или конкурирующая гипотеза?
3. Что называется статистическим критерием?
4. Дайте определение ошибкам первого и второго родов.
5. Что такое критическая область?
6. Какова основная идея критерия χ^2- Пирсона проверки гипотез?