МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ

ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ БЮДЖЕТНОЕ ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ «РОССИЙСКИЙ ГОСУДАРСТВЕННЫЙ ПЕДАГОГИЧЕСКИЙ УНИВЕРСИТЕТ им. А. И. ГЕРЦЕНА

Институт информационных технологий и технологического образования

Кафедра компьютерные технологии и электронного обучения

Основная профессиональная образовательная программа

Направление подготовки 09.03.01 Информатика и вычислительная техника

Направленность (профиль) «Технологии разработки программного обеспечения»

форма обучения – очная

ЛАБОРАТОРНАЯ РАБОТА №6

по дисциплине: «Анализ данных и основы Data science»

ПРОВЕРКА СТАТИСТИЧЕСКИХ ГИПОТЕЗ

Руководитель:

кандидат педагогических наук, доцент,

Светлана Викторовна Гончарова

Автор работы студент 2 курса
1 группы 1 подгруппы
Чирцов Тимофей Александрович

Цель работы: проверить статистическую гипотезу о нормальном законе распределения данных, приведенных в решаемой задаче.

Санкт-Петербург
2023

Оборудование: ПК, табличный процессор Excel

<u>Задание 1</u>

Из нормальной генеральной совокупности с известной дисперсией $\sigma^2 = 3{,}2$ извлечена выборка объёма $n = 25$ и по ней найдена выборочная средняя $\bar{x}_в = 19{,}3$. Требуется на уровне значимости 0,01 проверить нулевую гипотезу $H_0 : a = 20$ против конкурирующей гипотезы $H_1 : a = 19$.

$H_0 : a = 20 \Rightarrow a_0 = 20$

$H_1 : a = 19 \Rightarrow a_1 = 19$

$a_0 > a_1$, следовательно, левосторонняя область

Определим критическое значение из соотношения:

$$\Phi(u_{кр}) = \frac{1 - 2\alpha}{2} = \frac{1 - 2 * 0{,}01}{2} = 0{,}49$$

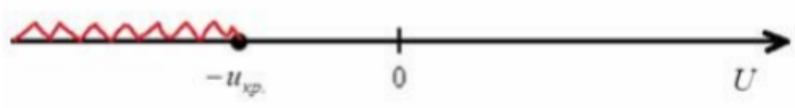
где α – уровень значимости

Далее по таблице значений функции Лапласа определяем критическое значение $u_{кр} \approx 2{,}33$

Для левосторонней области

Если $u < -u_{кр}$, то гипотеза H_0 на уровне α отвергается

Если $u > -u_{кр}$, то гипотеза H_0 на уровне α принимается



$$u_{набл} = \frac{(\bar{x}_в - a_0)\sqrt{n}}{\delta} = \frac{(19{,}3 - 20)\sqrt{25}}{\sqrt{3{,}2}} = -1{,}95656$$

$-1{,}95656 > -2{,}33$

Следовательно, $u_{набл} > -u_{кр}$

Значит, на уровне α = 0,01 гипотеза H_0 принимается

Задание 2

По результатам $n = 5$ измерений температуры в печи найдено $\bar{x}_в = 256°C$. Предполагается, что ошибка измерения есть нормальная случайная величина с $\sigma = 6°C$. Проверить на уровне значимости $\alpha = 0{,}05$ гипотезу $H_0: a = 250°C$ против конкурирующей гипотезы $H_1: a > 250°C$.

$H_0 : a = 250 \Rightarrow a_0 = 250$

$H_1 : a > 250$, следовательно, правосторонняя область

Определим критическое значение из соотношения:

$$\Phi(u_{кр}) = \frac{1 - 2\alpha}{2} = \frac{1 - 2 * 0{,}05}{2} = 0{,}45$$

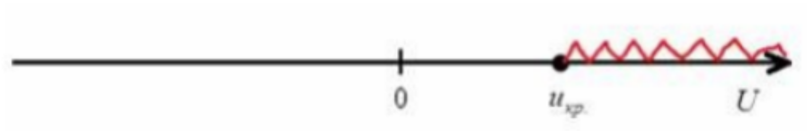
где α – уровень значимости

Далее по таблице значений функции Лапласа определяем критическое значение $u_{кр} \approx 1{,}645$

Для правосторонней области

Если $u < u_{кр}$, то гипотеза H_0 на уровне α принимается

Если $u > u_{кр}$, то гипотеза H_0 на уровне α отвергается



$$u_{набл} = \frac{(\overline{x_в} - a_0)\sqrt{n}}{\delta} = \frac{(256 - 250)\sqrt{5}}{6} = 2{,}23607$$

2,23607 > 1,645

Следовательно, $u_{набл} > u_{кр}$

Значит, на уровне α = 0,05 гипотеза H_0 отвергается

Задание 3

Задача Рез-ты исследований прочности на сжатие (СВ X) - 200 образцов бетона - представлены в виде сгруппированного стат. ряда.

Интервал прочности кг/см²	Среднее значение интервала, x_i	Частота, n_i
190 - 200	195	10
200 - 210	205	26
210 - 220	215	56
220 - 230	225	64
230 - 240	235	30
240 - 250	245	14

$\sum n_i = n = 200$

Проверить нулевую гипотезу о нормальном з-не распределения прочности образцов бетона на сжатие. Уровень значимости $\alpha = 0{,}001$.

Для проверки H0 найдём точечные оценки математического ожидания и среднего квадратичного отклонения нормального распределения случайной величины:

$$\bar{X} = \frac{\sum x_i n_i}{n}$$

$$\sigma = \sqrt{\frac{1}{n} \sum (x_i - \bar{X})^2 \cdot n_i} =$$

При проверке гипотезы о нормальном распределении генеральной совокупности сравниваются эмпирическая и теоретическая частоты. Для этого исключается статистика x^2 – Пирсона с $v = k\text{-}r\text{-}1$ степенями параметров.

Если $x^2_{расч} \geq x^2_{кр}$, то H_0 отвергается, и считается, что предположение о нормальном распределении не согласуется с данными

Вычислим теоретические вероятности p_i попадания СВ $X \rightarrow N$ (221;12,33) в частичные интервалы $[x_{i-1}; x_i)$:

$$U_i = \frac{x_i - \bar{X}}{\sigma}$$

$$\varphi(u_i)=\frac{1}{\sqrt{2\pi}}\int_0^{u_i} e^{-\frac{t^2}{2}}\,dt$$

Дальнейшие вычисления, необходимые для определения расчетного значения x^2, сделаем в табличном процессоре Excel:

	A	B	C	D	E	F	G	H	I	J	K	L	M
1		Интервалы прочности кг/см^2	Среднее значение интервала, x_i	Частота, ni	x_i-X̄	$(x_i$-X̄$)^2$	$(x_i$-X̄$)^2*n_i$	u_i	Нормированные интервалы $[u_i; u_{i+1}]$	$P_i = [\Phi(u_{i+1})-\Phi(ui)]$	nP_i	$(n_i-n*P_i)^2$	$(n_i-nP_i)^2/nP_i$
2		190-200	195	10	-26	676	6760	-2,51	(-∞; -1,70)	0,045	9	1	0,11
3		200-210	205	26	-16	256	6656	-1,70	[-1,70; -0,89]	0,142	28,4	5,76	0,20
4		210-220	215	56	-6	36	2016	-0,89	[-0,89; -0,08]	0,281	56,2	0,04	0,00
5		220-230	225	64	4	16	1024	-0,08	[-0,08; 0,73]	0,299	59,8	17,64	0,29
6		230-240	235	30	14	196	5880	0,73	[0,73; 1,54]	0,171	34,2	17,64	0,52
7		240-250	245	14	24	576	8064	1,54	[1,54; +∞)	0,062	12,4	2,56	0,21
8	Сумма	-	-	200	-6	1756	30400	-	-	-	200	-	1,33
9													
10	n	200											
11	X̄	221											
12	δ	12,328828	12,33										

В результате вычислений получили $x^2_{расч} = 1,35$

$v = k-r-1 = 6-2-1 = 3$

Следовательно, по таблице квантилей критическое значение будет равно:

$x^2_{кр} = 16,266$

То есть $x^2_{расч} < x^2_{кр}$ (1,35 < 16,266), то нет оснований для отклонения нулевой гипотезы о нормальном законе распределения прочности на сжатие с параметрами

$a = 221$ и $\delta^2 = 152$

Вывод по лабораторной работе: с помощью электронных таблиц нам удалось проверить статистическую гипотезу о нормальном законе распределения данных.