

**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 №4
по дисциплине: «Анализ данных и основы Data science»
НОРМАЛЬНОЕ РАСПРЕДЕЛЕНИЕ**

Руководитель:
кандидат педагогических наук, доцент,
Светлана Викторовна Гончарова

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

Санкт-Петербург
2022

Цель: проверить данные на соответствие нормальному закону распределения

Оборудование: Персональный компьютер, Excel

Задание 1

(3) Постановка задачи:

Рассчитать теоретические частоты для нормального распределения.

(4) Математические модели:

$$\bar{X} = \frac{\sum x_i \cdot f_i}{\sum f_i}$$

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

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

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

$$f_m = \varphi(t) \frac{Nd}{\sigma}$$

Где: N – объем совокупности (сумма всех эмпирических частот)
d – длина интервала (если вариационный ряд построен с неравными интервалами, то d будет меняться при переходе от одного интервала к другому)

(5) Результат выполненной работы:

Сумма затрат предприятий на производство, тыс. руб.	Количество предприятий, f_i	Середина интервала, x_i	$x_i * f_i$	\bar{x}	$(x_i - \bar{x})^2$	$(x_i - \bar{x})^2 * f_i$	σ	t_i	$\varphi(t)$	N	d	f_m
A	1	2						3	4			5
30-40	2	35	70	97,222	3871,605	7743,21	29,638	-2,10	0,0440	108	10	2
40-50	4	45	180		2727,16	10908,64		-1,76	0,0845			3
50-60	6	55	330		1782,716	10696,3		-1,42	0,1446			5
60-70	8	65	520		1038,272	8306,173		-1,09	0,2209			8
70-80	11	75	825		493,8272	5432,099		-0,75	0,3012			11
80-90	14	85	1190		149,3827	2091,358		-0,41	0,3664			13
90-100	15	95	1425		4,938272	74,07407		-0,07	0,3978			14
100-110	13	105	1365		60,49383	786,4198		0,26	0,3854			14
110-120	11	115	1265		316,0494	3476,543		0,60	0,3333			12
120-130	8	125	1000		771,6049	6172,84		0,94	0,2571			9
130-140	6	135	810		1427,16	8562,963		1,27	0,1771			6
140-150	5	145	725		2282,716	11413,58		1,61	0,1088			4
150-160	3	155	465		3338,272	10014,81		1,95	0,0597			2
160-170	2	165	330		4593,827	9187,654		2,29	0,0292			1
ИТОГ	108	-	-	-	-	-	-	-	-	-	-	106

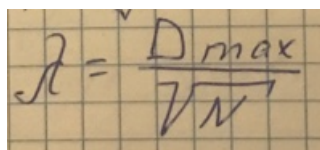
(6) Вывод по заданию 1: Видно, что вычисленные теоретические частоты достаточно близки к эмпирическим.

Задание 2

(3) Постановка задачи:

Определить является ли распределение (из Задания 1) нормальным. Использовать критерий Колмогорова.

(4) Математическая модель:



$$D = \frac{D_{\max}}{\sqrt{n}}$$

(5) Результат выполненной работы:

F_i	F_m	D_i	D_{\max}	λ
6	7	8		9
2	2	0	3	0,305143
6	5	1		
12	10	2		
20	18	2		
31	29	2		
45	42	3		
60	57	3		
73	71	2		
84	83	1		
92	92	0		
98	99	1		
103	103	0		
106	105	1		
108	106	2		
-	-	-	-	-

Вывод по всей лабораторной работе: С помощью электронных таблиц мы реализовали проверку данных на соответствие нормального закона распределения. Таким образом, делаем вывод, что распределение исследуемых данных достаточно близок к нормальному закону.