## ПРИЛОЖЕНИЕ 3

Проявляйте заботу о мухах, а слоны сами о себе позаботятся. Мерфология, Зақон заменяемости Питера

## Полигон и гистограмма частостей, эмпирическая функция распределения

В приложении представлен пример построения эмпирической функции распределения, полигона и гистограммы частостей. Указанные построения выполняются после того, как предварительно проведен отсев грубых погрешностей.

Пример ПЗ. Пусть, допустим, при помощи правила "ящик с усами" (см. Приложение 2) получен массив эмпирических данных, представленный на рис. ПЗ.1 соответствующими диапазонам ячеек для пяти ярусов листвы. Для удобства вычислений этим диапазонам присваиваются нижеследующие имена.

номер яруса	Диапазон	Имя
1	A4:H8	X_1
2	J4:Q8	X_2
3	A11:H15	X_3
4	J11:Q16	X_4
5	A18:H23	X_5
вся крона	A4:Q23	Х



1		A	В	С	D	E	F	G	Н	1	J	К	L	М	Ν	0	Ρ	Q	
	1	Разм	иери	пист	ьев	по я	руса	ам к	оонь	,									
I	2																		
	3	ЯРУС	1								ЯРУС	0 2							
	4	11,4	11,9	11,5	11,6	12,0	11,5	11,1	11,3			14,3	14,4	14,9	14,3	17,5	17,5	17,7	
	5	12,4	12,1	12,6	12,1	12,5	12,2	14,1	14,8		11,4	11,8	11,4	16,3	16,1	11,4	11,9	15,8	
	6				8,2	10,1	10,7	10,4	10,7		12,1	12,5	12,2	17,0	16,6	12,3		17,3	
	7			13,7	13,9	13,2	13,8	16,3	16,2		13,2	13,9	13,0	14,4	14,1	13,9	13,8	13,5	
	8	14,7	14,3	14,8	14,8	15,2	15,6	15,5	14,7		15,6	15,5	15,3	15,1	15,1	15,0	15,1	15,8	
	9																		
	10	ЯРУС	3								ЯРУС	РУС 4							
	11		12,6	12,1	12,5	12,2	9,1	9,3	9,3			16,5	16,9	16,9	16,1	17,0	11,1	11,2	
	12			14,8	14,8	14,1	14,8	14,7	14,4			10,2	10,2	10,5	16,8	16,1	16,8	16,5	
	13	19,9	11,9	11,5	17,5	18,0	17,9	17,9	17,3			12,9	12,7	12,8	12,5	12,1	14,9	14,2	
	14	16,5	18,8	18,4	18,5	18,9	18,5	20,8	20,4		9,0	13,5	13,7	13,5	18,5	18,9	20,7	20,8	
l	15	8,0	8,2	15,5	15,4	15,1	15,8	15,8	13,2		17,3	17,7	17,5	17,8	17,8	17,0	17,6	17,2	
	16										18,1	19,0	18,1	20,7	15,3	17,5	17,8	19,3	
	17	ЯРУС	5																
	18	11,8	11,5	16,8	16,1	16,8	16,8	16,3	16,3										
	19	12,6	12,0	12,4		18,9	18,5	18,6	16,6										
	20	13,1	13,9	13,0	17,7	17,7	17,1	17,1	16,7										
	21	14,8	14,4	14,3	14,6	14,5	14,4	14,4	16,3										
	22	15,1	15,3	15,5	15,5	15,4	15,4	15,4	16,2										
	23	19,8	19,1	19,8	17,2	17,4	17,7												
411	0.4																		

Рис. ПЗ.1. Скриншот расчетного листа MS Excel (исходные данные)

Последовательность обработки данных следующая.

- Вводится пояснительная информация (см. рис. ПЗ.2); в ячейку Р25 заносится величина уровня значимости.
- 2. В ячейках В25 и F25 формулами =МИН(Х) {=MIN(Х)} и =МАКС(Х) {=МАХ(Х)} определяются минимальное и максимальное значения по всей выборке, а в ячейке К25 формулой =СЧЁТ(Х) {=COUNT(Х)} вычисляется объем этой выборки.
- 3. В ячейке Е27 формулой =ОКРУГЛВНИЗ (1+LOG(K25;2);0) {=ROUNDDOWN (1 +LOG(K25;2); 0)}, реализующей соотношение Sturgess'а, определяется количество классов (интервалов) для выборки. Интервальный шаг вычисляется в ячейке Е28 формулой =(F25-B25)/E27.
- 4. Формируются границы классов: в ячейку А30 формулой =В25 заносится нижняя граница первого интервала, а в В30 формулой =А30+Е\$28 его верхняя граница. В ячейку А31 заносится нижняя граница второго интервала формулой =А30+Е\$28. Механизмом автозаполнения ячейка А31 тиражируется на диапазон А32:А37, а ячейка В30 на диапазон В31:В37.
- **5.** В ячейке C30 формулой =(A30+B30)/2 рассчитывается центр интервала, далее эта ячейка автозаполнением тиражируется на диапазон C31:C37.

	А	В	С	D	E	F	G	Н	Ι	J	K	L	М	Ν	0	Ρ	Q	R
24																		
25	мин=	8,0		N	акс=	20,8				N=	200				α=	0,05		
26																		
27	число	ов =	8,0						для в	сей е	выбо	ки						
28	шаг п	о инт	ервал	⊓y =	1,6						Мода							14,80
29	Хнач	Хкон	Хсред	ni	<b>p</b> *	F(x)					Медиа	ана						14,95
30	8,0	9,6	8,8	7	0,04	0,04					Средн	неари	фмет	ичесн	oe			14,88
31	9,6	11,2	10,4	10	0,05	0,09												
32	11,2	12,8	12,0	36	0,18	0,27					Вариа	цион	ный р	азма	Х			12,80
33	12,8	14,4	13,6	27	0,14	0,40					Средн	еква,	драти	чное	откло	нени	е	2,77
34	14,4	16,0	15,2	45	0,23	0,63					Диспе	рсия						7,65
35	16,0	17,6	16,8	40	0,20	0,83					Коэфс	рицие	ент ва	ариац	ии			18,6%
36	17,6	19,2	18,4	26	0,13	0,96												
37	19,2	20,8	20,0	9	0,05	1,00					Довер	ителі	ьный	инте	звал			0,383
38			N=	200														

Рис. ПЗ.2. Скриншот расчетного листа MS Excel (статистические показатели)

- 7. В ячейке E30 формулой =D30/D\$38 рассчитывается частость для первого интервала, далее эта ячейка автозаполнением тиражируется на диапазон E31:E37.

- 8. Для подсчета кумулятивной функции эмпирического распределения в ячейку F30 заносится частость первого интервала формулой =E30; в ячейку F31 заносится формула накопления =F30+E31, после чего данная ячейка тиражируется автозаполнением на диапазон F32:F37.
- 9. Для всей выборки по нижеследующим адресам рассчитываются основные статистические показатели.

наименование показателя	адрес	расчетная формула
Мода	R28	=МОДА(X) {=MODE(X)}
Медиана	R29	=МЕДИАНА(X) {=MEDIAN(X)}
Среднее арифметическое	R30	=CP3HAЧ(X) {=AVERAGE(X)}
Вариационный размах	R32	=F25-B25
Среднее квадратичное (стандартное) отклонение	R33	=CTAHДOTKЛOH(X) {=STDEV(X)}
Дисперсия	R34	=ДИСП(X) {=VAR(X)}
Коэффициент вариации.	R35	=R33/R30
Ожидаемый доверительный интер- вал для математического ожидания генеральной совокупности	R37	=ДОВЕРИТ(P25;R33;K25) {=CONFIDENCE(P25;R33;K25)}

10. По диапазонам значений аргумента C30:C37 и функции E30:E37 инструментом "диаграмма" строятся полигон (рис.П3.3) и гистограмма (рис.П3.4) частостей. По гистограмме для наглядности проводится линия тренда.



Рис. ПЗ.3. Скриншот расчетного листа MS Excel



- **11.** По диапазонам значений аргумента C30:C37 и функции F30:F37 инструментом "диаграмма" строится эмпирическая функция распределения (рис.П3.5).
- 12. В ячейке ОЗО (рис. ПЗ.6) формулой =КОРРЕЛ(В100:F100;В103:F103) {=CORREL (В100:F100;
  В103:F103)} рассчитывается коэффициент корреляции.

Поскольку это значение достаточно велико, т.е. зависимость между номером яруса и средним по ярусу размером листа существенна, то имеет смысл построить линейную регрессию, связывающую эти величины.



Рис. ПЗ.5. Скриншот расчетного листа MS Excel

13. Поясняющими надписями заполняются ячейки A100:G100, A101:A109. Подсчитываются размеры выборок по ярусам и полный объем.

В ячейку В101 вводится формула =СЧЁТ(Х\_1) {=COUNT(Х\_1)}, в С101 – =СЧЁТ(Х\_2) {=COUNT(X\_2)}, в D101 – =СЧЁТ(Х\_3) {=COUNT(X\_3)}, в E101 – =СЧЁТ(Х\_4) {=COUNT(X\_4)}, в F101 – =СЧЁТ(Х\_5) {=COUNT(X\_5)}, в G101 – =СУММ(В101:F101) {=SUM(B101:F101)}.

- 14. Определяются степени свободы формула =В101-1, внесенная в В102, автозаполнением тиражируется по диапазону С102:G102.
- 15. Подсчитываются размеры листьев по ярусам и полному объему: в ячейку B103 вводится формула =CP3HAЧ(X\_1) {=AVERAGE(X\_1)}, в C103 – = CP3HAЧ(X\_2) {=AVERAGE(X\_2)}, в D103 – = CP3HAЧ(X\_3) {=AVERAGE(X\_3)}, в E103 – =CP3HAЧ(X\_4) {=AVERAGE(X\_4)}, в F103 – =CP3HAЧ(X\_5) {=AVERAGE (X\_5)}, в G103 – =CP3HAЧ(X) {=AVERAGE(X)}.
- 16. Аналогичным образом определяются стандартные отклонения: в ячейку В104 вводится формула =СТАНДОТКЛОН(Х\_1) {=STDEV(X\_1)}, в С104 – =СТАНДОТКЛОН(Х\_2) {=STDEV(X\_2)}, в D104 – =СТАНДОТКЛОН (Х\_3) {=STDEV(X\_3)}, в Е104 – =СТАНДОТКЛОН(Х\_4) {=STDEV (X\_4)}, в F104 – =СТАНДОТКЛОН(Х\_5) {=STDEV(X\_5)}, в G104 – =СТАНДОТКЛОН(Х) {=STDEV(X)}.
- 17. Определяются размеры и границы доверительных интервалов:
  - формула =ДОВЕРИТ(\$P25; B104; B101) {=CONFIDENCE(\$P25; B104; B101)}, внесенная в В105, автозаполнением тиражируется по диапазону C105:G105;
  - формула =B103-B105, внесенная в B106, тиражируется на C106:G106;
  - формула =B103+B105, внесенная в B107, автозаполнением тиражируется по диапазону C107:G107;

	Α	В	С	D	E	F	G	Н	I	J	K	L	Μ	Ν	0		
99																	
100	Ярус	1	2	3	4	5	Σ			коэф	фициен	нт кор	реляц	ции=	0,9		
101	n=	35	38	37	45	45	200										
102	df=	34	37	36	44	44	199			урав	нение	пиней	йной ј	регре	ссии		
103	х ср =	12,9	14,4	15,0	15,8	15,8	14,9			x регр = m*Яр + b							
104	σ=	1,95	1,87	3,51	3,06	2,10	2,84						Яр - н	юмер	яруса		
105	d=	0,6	0,6	1,1	0,9	0,6	0,4			<i>m</i> =	b=						
106	x-d=	12,3	13,9	13,9	14,9	15,2	14,5			0,71	12,66						
107	x+d=	13,6	15,0	16,1	16,7	16,4	15,3										
108																	
109	х регр	13,4	14,1	14,8	15,5	16,2											

Рис. ПЗ.6. Скриншот расчетного листа MS Excel

- 18. Определяются коэффициенты уравнения линейной регрессии: в ячейку J106 вводится формула =ЛИНЕЙН(В103:F103; B100:F100; ИСТИНА; ЛОЖЬ) {=LINEST(B103:F103; B100:F100; TRUE(); FALSE())}, копируется в K106; обе ячейки выделяются и "активируются" через F2 и Ctrl+Shift+Enter. В OpenOfficeCalc выделяются обе ячейки, вводится формула и через Ctrl+Shift+Enter активируется.
- 19. Из уравнения регрессии подсчитываются значения размеров листьев по ярусам: формула =\$J106\*B100+\$K106, внесенная в B109, автозаполнением тиражируется по диапазону C109:F109.

20. По строкам данных B100:F100, B103:F103, B106:F106, B107:F107, B109:F109 инструментом "диаграмма" строится обобщающий график, представленный на рис. ПЗ.7.

