Шашков В.Б.

НЕКОТОРЫЕ РЕЗУЛЬТАТЫ КОМПЬЮТЕРНЫХ ЭКСПЕРИМЕНТОВ, МОДЕЛИРУЮЩИХ ПРОЦЕДУРЫ ОЦЕНИВАНИЯ ПАРАМЕТРОВ РАСПРЕДЕЛЕНИЯ

Разработана программная модель процедур статистического оценивания параметров распределения, включающая оригинальные уравнения для критических точек t- и χ^2-распределений. Получены уравнения и оценена сила стохастической связи между точностью выборочных оценок Mx^* и Dx^*, объемом выборок и генеральной дисперсией. Получены количественные показатели эффективности интервальных оценок, построенных по усредненным данным нескольких выборок.

Разработана программная модель выполнения статистических процедур, которая позволяет:

– создавать большие (например, по десять тысяч элементов) нормально распределенные массивы, моделирующие генеральные совокупности с заданными значениями математического ожидания и дисперсии;

– извлекать из этих совокупностей заданное количество выборок заданного объема (например, сто выборок объемом по сто элементов) с получением точечных и интервальных оценок параметров распределения при вероятностях 0,95; 0,90 и 0,80. При этом оценки по выборкам могут быть получены как «индивидуальные», так и «общие», т. е. усредненные по всей совокупности выборок;

– проводить массовые статистические эксперименты с целью получения экспериментальных данных, количество которых достаточно для построения соответствующих эмпирических формул, отражающих стохастические или функциональные зависимости между величинами – объектами эксперимента.

Первой задачей, возникающей при создании такой программной модели, является необходимость представления в ней материала статистических таблиц, содержащих критические точки распределения Стьюдента, χ^2-распределения Пирсона и т. п. В некоторых случаях содержание таких таблиц в подпрограммы вводится непосредственно (см., например, [1], с. 338). Более удобной и компактной формой является представление этих данных в аналитическом виде. С этой целью при выполнении настоящей работы были выведены эмпирические уравнения, воспроизводящие табличные данные с высокой точностью.

Использовались два вида эмпирических формул. Для нахождения уравнений нелинейной парной зависимости существует метод подбора наиболее подходящей формулы из набора наиболее часто применяемых функций типа

$$y = b_0 \times b_1^x \text{ или } y = \frac{x}{b_0 + b_1 \times x} \text{ и т. п.}$$

При этом функции линеаризуются, а коэффициенты b находят методом наименьших квадратов [2]. Такие уравнения будем называть парными зависимостями. Точность отражения экспериментальных данных их расчетными значениями будем оценивать значениями их коэффициентов корреляции.

Второй вид используемых уравнений – алгебраические степенные полиномы регрессии. Согласие между экспериментальными (или табличными) и расчетными данными при использовании полиномов будем оценивать по принятым показателям [3]. К ним относится остаточная дисперсия уравнения регрессии:

$$S_{os}^2 = \frac{\sum_{g=1}^{n} (y_g - yr_g)^2}{n-(k+1)}, \quad (1)$$

где y_g – экспериментальное или табличное значение функции, аппроксимируемое полиномом регрессии;

yr_g – рассчитанное по полиному значение функции;

g – номер строки таблицы данных;

n – количество строк в таблице;

(k+1) – количество коэффициентов регрессии в полиноме.

В качестве второго показателя качества полинома регрессии использовалось корреляционное отношенее /3/, которое по существу является коэффициентом функциональности рассматриваемого уравнения

$$KO = \frac{\sum_{g=1}^{n} (y_g - yr_g)^2}{\sum_{g=1}^{n} (y_g - \overline{y_g})^2}, \quad (2)$$

где $\overline{y_g}$ – среднее арифметическое вектора экспериментальных значений аппроксимируемой функции.

Следующим показателем качества полинома явилась интервальная оценка для математического ожидания расчетного значения функции $M\{yr_g\}$ (чем уже интервал, тем уравнение регрессии точнее [4]).

При построении уравнений для критических границ доверительных интервалов пользовались данными статистических таблиц источника [2].

Для критических точек распределения Стьюдента были найдены парные зависимости вида

$$t = \frac{\nu}{b_1 \cdot \nu - b_0}, \qquad (3)$$

где ν – число степеней свободы,

b_j – эмпирические коэффициенты.

Для всех принятых вероятностей коэффициент корреляции табличных и расчетных данных для этих уравнений был более 0,99999. Значения коэффициентов b_j составили:

– при p=0,95 $b_0 = -0{,}6130$; $b_1 = 0{,}5101$;

– при p=0,90 $b_0 = -0{,}5618$; $b_1 = 0{,}6079$;

– при p=0,80 $b_0 = -0{,}5151$: $b_1 = 0{,}7803$.

Эти уравнения и были введены в программную модель.

При определении критических точек χ^2-распределения Пирсона при значении числа степеней свободы ν до тридцати используют табулированные данные; если же $\nu > 30$, критическое значение χ^2 рассчитывают по уравнению Уилсона – Гилферти

$$\chi^2 = \nu\left(1 - \frac{2}{9\nu} \pm U_q\sqrt{\frac{2}{9\nu}}\right)^3, \qquad (4)$$

где U_q – квантиль нормального стандартного распределения при данном уровне значимости.

При проведении настоящей работы для двусторонних границ доверительных интервалов χ^2-распределения (при p = 0,95; 0,90 и 0,80) получены уравнения регрессии для чисел степеней свободы ν от четырех до ста. Уравнения выведены следующим образом. Значения критических границ χ^2-распределения при числе степеней свободы от 4 до 40 были заимствованы из работы [2]. Для значений ν от 35 до 100 они были рассчитаны по уравнению (4). Затем эти данные были объединены в единую задачу регрессии, где числа ν играли роль табличнозаданных аргументов, а значения χ^2 – таблично-заданных функций. Зависимость между ними искалась в виде полинома

$$y = b_0 + b_1\nu + b_2\nu^2 + \ldots\ldots\ldots + b_k\nu^k, \qquad (5)$$

где индекс при коэффициенте b показывает степень аргумента.

Для трех двусторонних интервальных оценок (при p = 0,95; 0,90 и 0,80) были получены шесть полиномов регрессии: три – для левой и три для правой границы. По своим качественным показателям каждое уравнение было «на 100% функцией»: остаточная дисперсия во всех случаях равнялась нулю, корреляционное отношение (коэффициент функциональности) составило единицу, а максимальная ширина интервала для математического ожидания расчетного значения χ^2 (т. е. величины $M\{\chi^2_{\text{расчет}}\}$) равнялась 0,007 – т. е. лежала в пределах погрешности расчетов.

Приводим значения коэффициентов b для всех шести уравнений (левые и правые границы) для возможного использования другими исследователями.

Индекс в скобках указывает степень аргумента ν:

b_lv80[0]=-1.38524144760049E+0000	b_pr80[0]= 2.22954056056187E+0000
b_lv80[1]= 5.28109666031924E-0001	b_pr80[1]= 1.47302074603431E+0000
b_lv80[2]= 1.21238672340382E-0002	b_pr80[2]=-1.21636478039306E-0002
b_lv80[3]=-2.55061285392433E-0004	b_pr80[3]= 2.55576622646636E-0004
b_lv80[4]= 3.19677527367557E-0006	b_pr80[4]=-3.19720082513422E-0006
b_lv80[5]=-2.12292825597135E-0008	b_pr80[5]= 2.11929940036504E-0008
b_lv80[6]= 5.73811784322987E-0011	b_pr80[6]=-5.71897286853806E-0011
b_lv90[0]=-1.31244621625228E+0000	b_pr90[0]= 3.50512875343338E+0000
b_lv90[1]= 4.03992189796554E-0001	b_pr90[1]= 1.60420421066738E+0000
b_lv90[2]= 1.51475995524208E-0002	b_pr90[2]=-1.55111792031164E-0002
b_lv90[3]=-3.18112800786086E-0004	b_pr90[3]= 3.26197400868367E-0004
b_lv90[4]= 3.99105805552052E-0006	b_pr90[4]=-4.08685529516145E-0006
b_lv90[5]=-2.65509656004030E-0008	b_pr90[5]= 2.71315541907908E-0008
b_lv90[6]= 7.18948434340249E-0011	b_pr90[6]=-7.33158733285371E-0011
b_lv95[0]=-1.17186016907363E+0000	b_pr95[0]= 4.78564764668408E+0000
b_lv95[1]= 3.00517778137419E-0001	b_pr95[1]= 1.71713127667135E+0000
b_lv95[2]= 1.76379571363725E-0002	b_pr95[2]=-1.84446730192178E-0002
b_lv95[3]=-3.70927740472204E-0004	b_pr95[3]= 3.90010160589593E-0004
b_lv95[4]= 4.66940739029392E-0006	b_pr95[4]=-4.91422881467724E-0006
b_lv95[5]=-3.11663587672026E-0008	b_pr95[5]= 3.27872658620007E-0008
b_lv95[6]= 8.46228740605203E-0011	b_pr95[6]=-8.89615303501841E-0011

Уравнения позволяют строить интервальные оценки по выборкам, содержащим до ста элементов.

Построение нормально распределенных массивов (играющих роль генеральных совокупностей) объемом в десять тысяч элементов проводилось по известной методике в соответствии с теоремой Ляпунова, т. е. случайные величины создавались как суммы слагаемых функции random. Соответствие распределения нормальному проверялось по значению математического ожидания и дисперсии нормированного массива (близость этих показателей нулю и единице), а также по значению коэффициентов асимметрии и эксцесса.

Объектом исследования явились закономерности, связывающие параметры генеральных совокупностей и их оценок.

Введем некоторые обозначения. Оцениваемый параметр теоретического распределения обозначим как θ, а его выборочную оценку как θ^*. Математическое ожидание и дисперсию генеральных совокупностей обозначим как MxG и DxG, а их оценки как MxV и DxV.

Проводили исследование зависимости разности $\Delta\theta = \theta - \theta^*$ от объема выборок и от генеральной дисперсии. Выборки разного объема, от пяти до ста элементов, извлекались случайным образом из генеральной нормально распределенной совокупности объемом десять тысяч элементов с математическим ожиданием 149,65 и дисперсией 2399,12. Образование выборок каждого объема дублировалось девять раз для сглаживания разбросов случайных результатов, перекрывающих влияние увеличения объема выборок. Данные эксперимента сводились в таблицу, где объем выборок играл роль аргумента, а разность $\Delta\theta$ являлась функцией.

Результаты эксперимента сведены в таблицу 1. Для одиннадцати групп выборок таблица экспериментальных данных содержала девяносто девять строк наблюдений (по девять наблюдений на каждую группу). Табличные функции (MxG – MxV) и (DxG – DxV) были приняты за показатели точности выборочных оценок.

В столбце 3 таблицы 1 из девяти параллельных наблюдений разности (MxG – MxV) по каждой группе выборок представлены два значения – минимальное и максимальное. В столбце 6 аналогичные данные приведены для разности (DxG – DxV). Эти наблюдения имели большой разброс значений, отражающих случайный характер величин, который часто перекрывал эффект увеличения объема выборки. Приведем максимальный пример такого «несоответствия» – разности (DxG – DxV) для выборок с объемом сорок (первая строка) и пятьдесят (вторая строка) элементов:

8,5 11,5 24,0 32,7 126,3 146,4 254,4 379,1 655,7
67,5 82,4 295,2 445,2 453,4 621,1 650,1 832,9 1112,8.

Как видим, выборки меньшего объема в данном случае имеют более хорошие показатели точности оценок. Этот пример наглядно демонстрирует, что даже девятикратное дублирование наблюдений не исключает «перевеса» фактора случайного разброса значений.

Приняв данные столбца два таблицы 1 за аргументы, а столбцов три и шесть за функции, нашли для зависимостей $\Delta Mx = f(n)$ и $\Delta Dx = g(n)$ по два вида уравнений: наилучшую парную зависимость из набора разных функций и полином регрессии с максимальными показателями качества – остаточной дисперсии уравнения $S^2_{ост}$ и корреляционного отношения КО.

Для разности ΔDx = (DxG – DxV) получена парная зависимость вида

$$\Delta Dx = 158{,}077+6781{,}484/n \qquad (6)$$

при коэффициенте корреляции, равном 0,765.

Наилучший из ряда найденных полином регрессии имеет довольно сложную форму (см. табл.)

$$\Delta Dx=b_0+b_1n+b_2\,n^2+b_3\,n^3+b_4\cdot$$
$$\cdot n^9+b_5\,n^{10}+b_6\,n^{11}+b_7\,n^{12}+b_8\,n^{13} \qquad (7)$$

при корреляционном отношении КО, равном 0,790.

Для разности (MxG – MxV) уравнения имели вид:

$$\Delta Mx = 23{,}799 - 4{,}583\cdot\ln(n) \qquad (8)$$

при коэффициенте корреляции, равном -0,607, и

$$\Delta Mx = b_0+b_1n+b_2\,n^2+b_3\,n^3+b_4\,n^4 \qquad (9)$$

при корреляционном отношении КО, равном 0,610.

Значения коэффициентов корреляции и корреляционного отношения показывают, что полученные уравнения довольно далеки от функциональной связи и отражают стохастическую связь между величинами со случайным разбросом значений. С другой стороны,

Таблица 1. Зависимость разности $\Delta\theta = \theta - \theta^*$ от объема выборок

№ пп	Объем выборок n	ΔMx экспериментальная	ΔMx по урав. (8)	ΔMx по урав. (9)	ΔDx экспериментальная	ΔDx по урав. (6)	ΔDx по урав. (7)
1	5	8,27 36,92	16,42	17,06	470,58 2299,91	1514,37	1602,04
2	10	1,71 23,63	13,25	13,62	10,00 951,39	836,23	607,05
3	20	3,58 23,42	10,07	9,22	18,65 1277,25	497,15	599,80
4	30	0,21 20,66	8,21	7,09	12,39 655,94	384,13	323,77
5	40	1,07 15,66	6,89	6,20	8,48 655,73	327,61	306,18
6	50	0,33 9,12	5,87	5,80	67,49 1112,84	293,71	376,56
7	60	0,22 12,07	5,04	5,39	9,17 676,46	271,10	374,08
8	70	0,21 9,87	4,33	4,77	2,71 781,68	254,95	293,25
9	80	0,05 6,43	3,72	3,96	30,78 450,15	242,85	239,58
10	90	0,99 8,59	3,18	3,27	56,44 324,05	233,43	183,90
11	100	0,02 6,03	2,69	3,29	0,92 471,41	225,89	175,23

они достаточно убедительно свидетельствуют о наличии несомненной количественной зависимости между точностью оценки $\Delta\theta$ и объемом выборки. При этом показатели этой связи объективны и устойчивы, так как совпадают у обоих видов полученных уравнений, т. е. и у парных зависимостей, и у полиномов регрессии.

Результаты расчета по полученым уравнениям представлены в таблице 1. Они показывают, что полиномы более «гибко» описывают случайный разброс данных, тогда как парные зависимости этот разброс сглаживают соответственно расчетной кривой элементарной функции.

Описанный выше эксперимент провели повторно, дублируя наблюдения по данной группе выборок не девять раз, а двукратно. Были получены уравнения такого же вида (при других значениях эмпирических коэффициентов). Показатели силы стохастической связи при этом составили соответственно: для ΔMx – 0,745 и 0,789; а для ΔDx 0,832 и 0,850. Как видим, результаты обоих экспериментов близки и отражают стохастический характер зависимостей.

При исследовании зависимости $\Delta\theta = f(DxG)$ были построены девятнадцать генеральных нормально распределенных совокупностей с разными дисперсиями DxG: от 100 до 10000. Математические ожидания у всех массивов были одинаковы и составили величины около 150 с небольшим разбросом, неизбежным при приближенном разыгрывании нормальной случайной величины. Из каждой совокупности последовательно извлекались по три выборки. Результаты эксперимента сведены в таблице 2.

Для зависимости $\Delta Dx = f(DxG)$ получено уравнение вида

$$\Delta Dx = 0{,}488 \cdot DxG^{0{,}888} \qquad (10)$$

при коэффициенте корреляции, равном 0,816, и полином вида

$$\Delta Dx = b_0 + b_1 DxG + b_2 DxG^2 + b_3 DxG^3 + b_4 DxG^4 \qquad (11)$$

при корреляционном отношении КО, равном 0,696.

Для разности $\Delta Mx = f(DxG)$ получена парная зависимость вида

$$\Delta Mx = 0{,}116 \cdot DxG^{0{,}562} \qquad (12)$$

при коэффициенте корреляции, равном 0,708, и полином вида

$$\Delta Mx = b_0 + b_1 DxG + b_2 DxG^2 + b_3 DxG^3 + b_4 DxG^9 + b_5 DxG^{10} + b_6 DxG^{11} \qquad (13)$$

при корреляционном отношении КО, равном 0,666.

Таким образом, характеристики стохастической связи для зависимостей $\Delta\theta = f(n)$ и $\Delta\theta = f(DxG)$ в отношении количественных показателей близки и устойчивы.

Все вышеизложенное посвящено вопросам точности точечных оценок. Но отдельная – единичная – точечная оценка не несет никакой информации о степени своей достоверности. Считается, что положение исправляет интервальная оценка, построенная на основании этой точечной:

$$\bar{x} - U_{\frac{q}{2}} \frac{\sigma_x}{\sqrt{n}} < Mx \le \bar{x} + U_{\frac{q}{2}} \frac{\sigma_x}{\sqrt{n}}, \qquad (14)$$

где U – квантиль нормального распределения стандартной формы x;

q – уровень значимости.

Утверждается, что «…если многократно извлекать выборки объемом N и с помощью не-

Таблица 2. Зависимость разности $\Delta\theta$ от генеральной дисперсии

№ пп	Ген. дисп. DxG	ΔMx экспериментальная	ΔMx по урав. (12)	ΔMx по урав. (13)	ΔDx экспериментальная	ΔDx по урав. (10)	ΔDx по урав. (11)
1	100,1	0,56 3,11	1,55	1,03	25,53 59,46	22,37	-14,94
2	200,0	1,02 3,40	2,28	2,31	17,47 62,20	53,84	32,67
3	400,6	1,98 3,36	3,37	4,28	17,95 225,05	99,74	119,94
4	610,2	2,46 6,50	4,27	5,70	65,65 205,05	144,91	199,06
5	807,0	4,37 15,12	5,00	6,63	132,25 343,66	185,72	263,39
6	995,4	3,36 13,07	5,62	7,27	178,28 302,67	223,75	317,12
7	1195,0	6,18 10,15	6,23	7,79	116,38 487,33	263,16	366,71
8	1383,8	2,37 15,66	6,77	8,20	112,64 689,41	299,75	407,71
9	1605,3	3,60 13,38	7,35	8,67	194,49 751,40	341,98	449,64
10	1824,9	3,01 16,54	7,90	9,18	690,49 1007,06	383,21	485,98
11	2000,8	5,11 11,72	8,32	9,66	201,20 602,59	415,83	512,19
12	2203,0	3,99 22,05	8,79	10,29	306,24 822,05	459,92	539,96
13	2399,1	7,96 11,54	9,22	10,79	436,33 1669,39	484,54	565,31
14	2680,5	2,27 17,34	9,81	12,13	201,88 439,18	539,06	600,44
15	2821,7	2,50 23,39	10,10	12,74	158,52 755,56	564,22	618,07
16	2986,1	10,34 18,27	10,42	13,44	203,66 1662,31	593,31	639,03
17	3629,1	13,32 16,51	11,63	15,44	236,38 992,27	705,44	731,56
18	6474,2	9,20 24,65	16,10	16,50	714,74 3205,70	1179,27	1598,23
19	10155,6	6,40 25,83	20,74	18,44	811,52 7018,21	1758,63	3251,98

равенства (14) определять доверительный интервал, то в среднем 100·p% построенных таким способом интервалов содержит истинное значение Mx» [4]. Однако в применении к единичной выборке это положение не увеличивает нашего знания о точности оценки. Если случайная величина $\bar{x}$ в неравенстве (14) есть среднее арифметическое ОДНОЙ выборки, то построение интервала с центром $\bar{x}$ ничего не скажет нам о степени близости оценки θ^* к оцениваемому параметру θ. И если оценка θ^* ошибочна, то построение интервала на это не укажет.

Проанализируем данные, приведенные в таблице 3. Из генеральной совокупности объемом десять тысяч элементов и с параметрами MxG = 149,56 и DxG = 3629,10 извлекли четыре выборки по десять элементов. По выборочным данным построили доверительные интервалы для обоих параметров распределения при вероятностях 0,80; 0,90 и 0,95. Как следует из результатов этого эксперимента, интервальные оценки для DxG на шестой, восьмой, девятой и двенадцатой стороках таблицы 3 не накрыли оцениваемый параметр DxG, имеющий значение 3629,10. Эти данные выделены полужирным шрифтом.

Были вычислены средние значения оценок MxV и DxV для всех четырех выборок. По ним при вероятности p = 0,90 были построены доверительные интервалы. Точечные оценки DxV выборок №2 и №3 не попали в интервал, построенный по среднему значению DxV для четырех выборок.

Результаты эксперимента, представленные в таблице 3, показывают, что информационная ценность интервальных оценок резко возрастает, если они построены хотя бы по небольшому количеству усредненных данных. Такая интервальная оценка будет единым обобщенным «измерителем пригодности» для всех выборок из данной генеральной совокупности. В этом случае интервальная оценка и будет работать в полном соответствии с цитатой. С целью практической количественной оценки положения, высказанного в данной цитате, провели эксперимент, результаты которого представлены в таблицах 4 и 5.

Таблица 3. Результаты интервального оценивания по отдельным и усредненным выборочным данным

№ выб.	Оценка MxV	Оценка DxV	Вер. P	Интервальные оценки			
				Для MxG		Для DxG	
				Лев. гр.	Пр. гр.	Лев. гр.	Пр. гр.
1	169,80	5012,59	0,95	119,13	220,46	2374,41	16586,33
			0,90	128,76	210,83	2669,57	13494,09
			0,80	138,83	200,76	3075,62	10783,40
2	157,50	**1541,96**	0,95	129,39	185,60	730,41	5102,24
			0,90	134,74	180,26	821,20	4151,01
			0,80	140,32	174,67	**946,11**	**3317,16**
3	146,09	**1323,97**	0,95	120,05	172,13	627,15	4380,92
			0,90	125,00	167,18	**705,11**	**3564,17**
			0,80	130,18	162,01	**812,36**	**2848,20**
4	127,81	6648,73	0,95	69,45	186,16	3149,43	22000,20
			0,90	80,54	175,07	3540,93	17898,63
			0,80	92,14	163,47	**4079,52**	**14303,15**
Ср.	**150,30***	**3631,81***	**0,90**	**115,37***	**185,23***	**1934,21***	**9777,53***

Было образовано пятнадцать нормально распределенных генеральных совокупностей. Математическое ожидание у всех массивов составило величину, близкую к 150, генеральная дисперсия возрастала от ста до 10000. Из каждой совокупности извлекали последовательно по сто выборок, объемом сто элементов каждая. Для всей совокупности выборок рассчитывались средние значения выборочных средних $\bar{\bar{x}}$ (т. е. среднее из средних) и среднее из выборочных дисперсий. При таком количестве выборок и их объемах эти усредненные оценки θ^* практически не отличались от оцениваемых параметров θ. По всем ста выборкам фиксировали размах выборочных показателей θ^*_{min} и θ^*_{max} (чтобы иметь возможность сравнить его с шириной доверительных интервалов). Наблюдение по ста выборкам из данного массива проводили дважды. По усредненным выборочным характеристикам $\overline{MxV}$ и $\overline{DxV}$ строили интервальные оценки и фиксировали количество выборок, характеристики которых не попадали в доверительные интервалы. Эти данные представлены в таблицах 4 и 5. Как следует из анализа данных, несмотря на существенный случайный разброс значений отдельных наблюдений, формулировка [4] «…то в среднем 100·p% построенных таким способом интервалов…» полностью подтвердилась.

Результаты настоящей работы позволяют сделать следующие выводы.

1. С помощью методов линеаризации элементарных функций в сочетании с методом наименьших квадратов получены уравнения парной связи для критических точек t-распределения Стьюдента. Для соответствующих показателей χ^2-распределения Пирсона получены полиномы регрессии шестой степени. Эти зависимости получены в виде отдельных уравнений для вероятностей 0,80; 0,90 и 0,95 и позволяют работать с выборками объемом до ста элементов. Полученные уравнения функциональны и отражают данные статистических таблиц с высокой точностью.

Таблица 4. Количество выборок, отсеянных по интервальным оценкам MxG при вероятностях 0,95 и 0,90

DxG массива	Среднее значение MxV	Размах MxV max-min	Интервальные оценки			
			P=0,95		P=0,90	
			Ширина интервала	Отсеяно Выборок	Ширина интервала	Отсеяно выборок
101,1	149,83	5,68	4,06	3	3,39	12
	150,21	4,60	3,93	4	3,29	10
200,0	150,07	7,16	5,64	4	4,71	11
	149,99	7,01	5,62	3	4,90	10
400,6	149,80	10,06	7,91	5	6,62	10
	149,90	6,66	7,86	0	6,57	1
610,2	149,64	12,66	9,77	8	8,17	16
	149,29	9,92	9,85	2	8,24	4
807,0	149,63	12,71	11,20	4	9,37	10
	149,49	14,60	11,29	4	9,44	9
995,4	149,96	15,32	12,67	6	10,56	10
	150,15	13,10	12,56	1	10,51	5
1195,0	150,45	18,37	13,81	5	11,55	7
	149,92	17,17	13,70	5	11,46	11
1605,3	149,85	20,31	15,96	7	13,35	9
	150,08	21,06	15,79	9	13,20	16
2000,8	150,72	20,71	17,60	4	14,72	9
	150,67	25,02	17,33	4	14,83	8
2399,1	150,07	26,64	19,40	6	16,23	11
	150,21	23,27	19,39	1	16,22	8
2821,7	150,61	38,17	20,89	5	17,47	8
	149,30	30,58	21,26	5	17,79	10
2986,1	149,43	34,75	21,89	5	18,31	11
	149,35	29,41	21,58	8	18,05	15
3629,1	150,79	35,01	23,84	5	19,94	10
	148,99	29,17	24,08	6	20,14	12
6474,2	153,09	39,66	31,92	4	26,70	6
	152,11	49,45	31,65	3	26,48	11
10155,6	148,78	44,37	40,13	6	33,56	10
	146,81	56,59	40,02	5	33,47	11
Итого в среднем отсеяно выборок				**4,57%**	-	**9,70%**

Таблица 5. Количество выборок, отсеянных по интервальным оценкам DxG при вероятностях 0,95 и 0,90

DxG массива	Среднее значение DxV	Размах DxV max-min	Интервальные оценки			
			P=0,95		P=0,90	
			Ширина интервала	Отсеяно выборок	Ширина интервала	Отсеяно выборок
101,1	104,43	72,24	58,61	4	47,92	9
	97,90	74,20	56,65	7	47,13	12
200,0	201,52	131,74	116,59	4	97,01	6
	200,69	135,68	116,11	4	96,61	8
400,6	397,39	347,03	229,82	4	191,31	8
	391,59	303,69	226,57	5	188,52	8
610,2	605,62	501,45	350,40	3	291,56	13
	616,29	479,04	356,57	7	296,69	13
807,0	796,15	494,36	460,64	7	383,28	11
	808,62	629,15	467,85	10	389,27	15
995,4	1011,19	769,91	585,06	6	486,81	7
	1001,32	811,55	579,35	5	482,05	11
1195,0	1210,60	831,62	700,43	8	582,80	16
	1191,40	741,23	689,32	6	573,56	13
1605,3	1615,80	1194,98	934,87	7	777,88	14
	1581,20	1031,44	914,85	9	761,22	18
2000,8	1966,27	1094,84	1137,65	7	946,60	11
	1995,36	1445,81	1154,48	11	960,60	17
2399,1	2388,63	1991,49	1382,02	7	1149,93	11
	2386,19	1802,21	1380,61	4	1148,76	13
2821,7	2767,88	2071,83	1601,45	9	1332,51	16
	2869,28	2180,89	1660,11	4	1381,32	10
2986,1	3040,42	1720,81	1759,13	2	1463,72	5
	2954,52	2201,55	1709,43	8	1422,36	13
3629,1	3606,74	2336,63	2086,80	3	1736,35	12
	3678,05	2684,86	2128,06	5	1770,68	9
6474,2	6465,54	4365,63	3740,85	4	3112,63	7
	6357,54	3435,24	3678,36	2	3060,64	4
10155,6	10218,14	8126,35	5912,04	7	4919,21	13
	10161,89	6648,35	5879,49	4	4892,13	12
Итого в среднем отсеяно выборок				**5,76%**	-	**11,17%**

2. Построены уравнения, отражающие стохастическую связь:

– между точностью статистических оценок в виде разности $(\theta - \theta^{*})$ и объемом выборок;

– между точностью статистических оценок в виде разности $(\theta - \theta^{*})$ и значением генеральной дисперсии.

Получены количественные характеристики силы такой связи в виде показателей степени функциональности уравнения. Ряд экспериментов как с одинаковыми, так и разными условиями проведения показал, что степень функциональности стохастических зависимостей $(\theta - \theta^{*}) = f(n)$ и $(\theta - \theta^{*}) = g(DxG)$ устойчиво держится в пределах от >0,60 до 0,85.

3. Показано, что информативность и «отсеивающая эффективность» интервальных оценок обеспечивается только при построении их по усредненным показателям нескольких выборок, извлеченных из данной генеральной совокупности.

4. При проведении массовых статистических экспериментов получены количественные показатели отсева выборок по интервальным оценкам в зависимости от принятой вероятности.

Список использованной литературы:

1. Плис А.И., Сливина Н.А. Лабораторный практикум по высшей математике. Учебное пособие. – М.: Высшая школа, 1994. – 416 с., ил.
2. Львовский Е.Н. Статистические методы построения эмпирических формул. – М.: Высшая школа, 1988. – 239 с., ил.
3. Пустыльник Е.И. Статистические методы анализа и обработки наблюдений. – М.: Наука, 1968. – 288 с., ил.
4. Бородюк В.П., Вощинин А.П., Иванов А.З. и др. Статистические методы в инженерных исследованиях. – М.: Высшая школа, 1983. – 216 с., ил.