Репина Е.Г.
Самарская государственная экономическая академия

ПРОБЛЕМА ИСПОЛЬЗОВАНИЯ КРИТЕРИЯ ГРАББСА НА ВЫБРОС ПРИ ЭКСПОНЕНЦИАЛЬНОМ ЗАКОНЕ РАСПРЕДЕЛЕНИЯ И ЗАКОНЕ РАСПРЕДЕЛЕНИЯ ЛАПЛАСА

В статье показано использование статистического критерия Граббса для проверки на аномальность наблюдений, принадлежащих выборкам, имеющим экспоненциальный закон распределения и закон распределения Лапласа.

Многие статистики в своих работах [1, 2] проявляют интерес к разработке робастных методов оценки экономической информации. Робастное оценивание незаменимо при обработке многомерной статистической информации, так как эмпирический метод выявления выбросов в исследуемой совокупности, не подтвержденный теоретическими выкладками, может привести к отбрасыванию слишком большого количества наблюдений.

При анализе экономического массива данных, как правило, считают, что показатели экономической деятельности предприятия подчиняются нормальному закону распределения. Практика же показывает, что совокупности значений многих экономических показателей не подчиняются теоретическому нормальному закону. Наблюдаются отклонения, как односторонние, так и двухсторонние, когда «хвосты» дифференциального закона оказываются более тяжелыми, чем можно было предположить, ориентируясь на данные таблиц нормального распределения.

В данной работе мы хотим показать применение робастного оценивания совокупностей, имеющих закон распределения, отличный от нормального, а именно использование критерия Граббса при отклонении наблюдаемого закона распределения от нормального закона.

Пусть проведена случайная повторная выборка объемом *n*: $X_1, X_2, ..., X_n$ и данные наблюдений упорядочены по возрастанию: $X_1 < X_2 < ... < X_n$.

Предположим, что случайная величина *X* имеет показательный (экспоненциальный) закон распределения. В этом случае плотность распределения (дифференциальная функция), как хорошо известно [3], имеет вид:

$$f(x) = \begin{cases} 0, x < 0, \\ \lambda e^{-\lambda e}, x \geq 0, \lambda > 0. \end{cases} \quad (1)$$

Тогда основная гипотеза H_0 заключается в том, что все данные наблюдений $X_1, X_2, ..., X_n$ принадлежат одной генеральной совокупности с показательным законом распределения.

Пусть осуществляется проверка на выброс одного (максимального) значения. В этом случае конкурирующая гипотеза H_1 состоит в том, что все X_j $(1 \leq j \leq n-1)$ имеют показательное распределение с показателем λ, а X_n подчиняется некоторому другому закону.

Если $X_n < d_{kp}$, то принимается гипотеза H_0, в противном случае – гипотеза H_1. Условие недопущения ошибки первого рода имеет вид $P(X_1 < d_{kp}, и\ X_2 < d_{kp}, и\ ..., и\ X_n < d_{kp}) = 1 - \alpha$, если гипотеза H_0 верна.

Поскольку выборка повторная, события вида $X_j < d_{kp}$ $j = \overline{1,n}$ можно считать независимыми. Тогда

$$P(X_1 < d_{kp}, и\ X_2 < d_{kp}, и\ ..., и\ X_n < d_{kp}) =$$
$$= P^n_{H_0}(X_j < d_{kp}),\ j = \overline{1,n}. \quad (2)$$

Поэтому

$$P^n_{H_0}(X_j < d_{kp}) = \sqrt[n]{1-\alpha} \quad (3)$$

или

$$F(d_{kp}) = \sqrt[n]{1-\alpha},$$

где $F(x) = 1 - e^{-\lambda x}$ – интегральная функция случайной величины *X*, распределенной по показательному закону с параметром λ.

В результате получим:

$$1 - e^{-\lambda d_{kp}} = \sqrt[n]{1-\alpha},$$

$$e^{-\lambda d_{kp}} = 1 - \sqrt[n]{1-\alpha},$$

$$d_{kp} = -\frac{1}{\lambda} ln(1 - \sqrt[n]{1-\alpha}).$$

Если параметр λ неизвестен, то можно найти его оценку $\hat{\lambda}$ по выборке: $\hat{\lambda} = \frac{1}{\bar{x}_в}$, а затем оценку $\hat{d}_{kp}$ по формуле $\hat{d}_{kp} = -\bar{x}_в\, ln(1 - \sqrt[n]{1-\alpha})$. С другой стороны, события

$X_j < d_{kp}$ и $\frac{X_j - M(X)}{\sigma(X)} < \frac{d_{kp} - M(X)}{\sigma(X)}$ – равновероятны. Поэтому можно записать

$$P_{H_0}\left(\frac{X_j - M(X)}{\sigma(X)} < \frac{d_{kp} - M(X)}{\sigma(X)}\right) = \sqrt[n]{1-\alpha}.$$

Обозначим $\frac{X_j - M(X)}{\sigma(X)} = G$, $\frac{d_{kp} - M(X)}{\sigma(X)} = G_{kp}$.

Если гипотеза H_0 верна, то $M(X) = \sigma(X) = \frac{1}{\lambda}$.

Тогда

$$G_{kp} = \frac{d_{kp} - \frac{1}{\lambda}}{\frac{1}{\lambda}} = \frac{-\frac{1}{\lambda} ln(1 - \sqrt[n]{1-\alpha}) - \frac{1}{\lambda}}{\frac{1}{\lambda}} =$$

$$= -(ln(1 - \sqrt[n]{1-\alpha}) + 1) = ln\left[\frac{1}{e(1 - \sqrt[n]{1-\alpha})}\right].$$

Итак,

$$G_{kp} = ln\left[\frac{1}{e(1 - \sqrt[n]{1-\alpha})}\right]. \qquad (4)$$

В качестве $G_{набл.}$ используем формулу

$$G_{набл.} \approx \frac{x_n - \bar{x}_в}{S}, \qquad (5)$$

где

$$\bar{x}_в = \frac{1}{n}\sum_{j=1}^{n} x_j, \quad S^2 = \frac{1}{n-1}\sum_{j=1}^{n}(x_j - \bar{x}_в).$$

Используя формулу (4), мы получили таблицу критических точек для $\alpha = 0{,}01$, $\alpha = 0{,}05$ и любого конечного *n* (табл.1).

Замечание. $G_{набл.}$ невозможно найти точно, так как неизвестна величина λ, и поэтому не можем вычислить $M(X) = \frac{1}{\lambda}$ и $\sigma(X) = \frac{1}{\lambda}$.

Если случайная величина X имеет распределение Лапласа (двустороннее экспоненциальное распределение), то, как известно [3], уравнение функции плотности имеет вид

$$f(x) = \frac{1}{2}\lambda e^{-\lambda|x|} \quad (-\infty < x < +\infty)$$

или

$$f(x) = \begin{cases} \frac{1}{2}\lambda e^{\lambda x}, x < 0, \\ \frac{1}{2}\lambda e^{-\lambda x}, x \ge 0. \end{cases} \qquad (6)$$

Интегральную функцию F(x) можно записать следующим образом

$$F(x) = \begin{cases} \frac{1}{2} e^{\lambda x}, если\ x < 0, \\ 1 - \frac{1}{2} e^{-\lambda x}, если\ x \ge 0. \end{cases}$$

Построим G_{kp}, проводя рассуждения и вычисления аналогично тому, как мы делаем это при экспоненциальном распределении.

Рассмотрим два случая:

1) $X_j \ge 0$ и 2) $X_j < 0$, $j = \overline{1,n}$.

<u>1 случай.</u> $X_j \ge 0$, $j = \overline{1,n}$. События $X_j < d_{kp}$, $d_{kp} > 0$ равновероятны, поэтому

$$P^n_{H_0}(X_j < d_{kp}) = 1 - \alpha,$$

$$P(X_j < d_{kp}) = \sqrt[n]{1-\alpha}$$

или

$$F(d_{kp}) = \sqrt[n]{1-\alpha},$$

где $F(x) = 1 - \frac{1}{2} e^{-\lambda x}$.

А тогда можно записать

$$1 - \frac{1}{2} e^{-\lambda d_{kp}} = \sqrt[n]{1-\alpha}.$$

Откуда легко получаем

$$d_{kp} = -\frac{1}{\lambda} ln\left[2(1 - \sqrt[n]{1-\alpha})\right].$$

<u>2 случай.</u> $X_j < 0$, $j = \overline{1,n}$. В этом случае имеем

$$P^n_{H_0}(X_j > C_{кр}) = 1 - \alpha, \ j = \overline{1,n}, C_{кр} < 0.$$

Или

$$\left[1 - F(C_{кр})\right]^n = 1 - \alpha,$$

или

$$F(C_{кр}) = 1 - \sqrt[n]{1-\alpha},$$

где $F(x) = \frac{1}{2} e^{\lambda x}$.

В результате получаем равенство

$$\frac{1}{2} e^{\lambda C_{кр}} = 1 - \sqrt[n]{1-\alpha}$$

Отсюда имеем

$$C_{кр} = \frac{1}{\lambda} ln\left[2(1 - \sqrt[n]{1-\alpha})\right]. \qquad (7)$$

Теперь выпишем

$$G_{кр} = \frac{d_{кр} - M(X)}{\sigma(x)}, d_{кр} > 0, X_j \ge 0, j = \overline{1,n}.$$

и

$$G_{кр} = \frac{M(X) - C_{кр}}{\sigma(x)}, C_{кр} < 0, X_j < 0, j = \overline{1,n}.$$

Если гипотеза H_0 верна (то есть все данные наблюдений принадлежат генеральной совокупности с распределением Лапласа), то

$$M(X) = 0, \sigma(X) = \frac{\sqrt{2}}{\lambda}.$$

А тогда

$$G_{кр} = \frac{d_{кр}}{\sqrt{2}}\lambda = -\frac{1}{\sqrt{2}} ln\left[2(1 - \sqrt[n]{1-\alpha})\right],$$

Таблица 1. Таблица критических значений критерия Граббса при экспоненциальном законе распределения исследуемой совокупности

Конечное число наблюдений, n	$G_{кр}$ при $\alpha = 0{,}01$	$G_{кр}$ при $\alpha = 0{,}05$	$G_{кр}$ при $\alpha = 0{,}1$
1	2	3	4
1	3,6052	1,9957	1,3026
2	4,2958	2,6761	1,9697
3	4,7004	3,0773	2,3665
4	4,9877	3,3629	2,6498
5	5,2106	3,5848	2,8703
6	5,3927	3,7662	3,0509
7	5,5468	3,9198	3,2038
8	5,6802	4,0528	3,3364
9	5,7979	4,1703	3,4534
10	5,9032	4,2753	3,5582
11	5,9985	4,3704	3,6530
12	6,0855	4,4572	3,7397
13	6,1655	4,5371	3,8194
14	6,2396	4,6111	3,8932
15	6,3085	4,6800	3,9619
16	6,3731	4,7444	4,0262
17	6,4337	4,8049	4,0867
18	6,4908	4,8620	4,1437
19	6,5449	4,9160	4,1976
20	6,5961	4,9672	4,2487
21	6,6449	5,0159	4,2974
22	6,6914	5,0624	4,3438
23	6,7359	5,1068	4,3882
24	6,7784	5,1493	4,4306
25	6,8192	5,1901	4,4713
26	6,8584	5,2293	4,5105
27	6,8962	5,2670	4,5482
28	6,9325	5,3033	4,5845
29	6,9676	5,3384	4,6195
30	7,0015	5,3722	4,6533
31	7,0343	5,4050	4,6861
32	7,0660	5,4367	4,7177
33	7,0968	5,4675	4,7485
34	7,1267	5,4973	4,7783
35	7,1556	5,5263	4,8072
36	7,1838	5,5544	4,8353
37	7,2112	5,5818	4,8627
38	7,2379	5,6085	4,8893
39	7,2638	5,6344	4,9153
40	7,2892	5,6597	4,9406
41	7,3138	5,6844	4,9652
42	7,3379	5,7085	4,9893
43	7,3615	5,7320	5,0128
44	7,3845	5,7550	5,0358
45	7,4069	5,7774	5,0582
46	7,4289	5,7994	5,0802
47	7,4504	5,8209	5,1016
48	7,4715	5,8419	5,1227
49	7,4921	5,8625	5,1433
50	7,5123	5,8827	5,1634
51	7,5321	5,9025	5,1832
52	7,5515	5,9219	5,2026
53	7,5705	5,9410	5,2217
54	7,5892	5,9597	5,2403
1	2	3	4
55	7,6076	5,9780	5,2587
56	7,6256	5,9960	5,2767
57	7,6433	6,0137	5,2943
58	7,6607	6,0311	5,3117
59	7,6778	6,0482	5,3288
60	7,6946	6,0650	5,3456
61	7,7111	6,0815	5,3621
62	7,7274	6,0977	5,3784
63	7,7434	6,1137	5,3943
64	7,7591	6,1295	5,4101
65	7,7746	6,1450	5,4256
66	7,7899	6,1602	5,4408
67	7,8049	6,1753	5,4558
68	7,8197	6,1901	5,4706
69	7,8343	6,2047	5,4852
70	7,8487	6,2191	5,4996
71	7,8629	6,2332	5,5138
72	7,8769	6,2472	5,5278
73	7,8907	6,2610	5,5415
74	7,9043	6,2746	5,5551
75	7,9177	6,2880	5,5686
76	7,9309	6,3013	5,5818
77	7,9440	6,3143	5,5949
78	7,9569	6,3272	5,6078
79	7,9697	6,3400	5,6205
80	7,9822	6,3525	5,6331
81	7,9947	6,3650	5,6455
82	8,0069	6,3772	5,6577
83	8,0191	6,3893	5,6698
84	8,0310	6,4013	5,6818
85	8,0429	6,4131	5,6936
86	8,0546	6,4248	5,7053
87	8,0661	6,4364	5,7169
88	8,0775	6,4478	5,7283
89	8,0888	6,4591	5,7396
90	8,1000	6,4703	5,7508
91	8,1111	6,4813	5,7618
92	8,1220	6,4923	5,7727
93	8,1328	6,5031	5,7835
94	8,1435	6,5138	5,7942
95	8,1541	6,5243	5,8048
96	8,1645	6,5348	5,8153
97	8,1749	6,5452	5,8256
98	8,1852	6,5554	5,8359
99	8,1953	6,5656	5,8460
100	8,2054	6,5756	5,8561
101	8,2153	6,5856	5,8660
102	8,2252	6,5954	5,8759
103	8,2349	6,6052	5,8856
104	8,2446	6,6148	5,8953
105	8,2542	6,6244	5,9048
106	8,2636	6,6339	5,9143
107	8,2730	6,6433	5,9237
108	8,2823	6,6526	5,9330
109	8,2915	6,6618	5,9422
110	8,3007	6,6709	5,9513
111	8,3097	6,6800	5,9604
112	8,3187	6,6889	5,9693

$$d_{кр} > 0\,,\, X_j \ge 0\,,\;\; j = \overline{1,n}$$

и

$$G_{кр} = -\frac{C_{кр}\cdot\lambda}{\sqrt{2}} = -\frac{1}{\sqrt{2}}\,ln\left[2(1-\sqrt[n]{1-\alpha}\,)\right],$$

где $C_{кр} < 0$ и $X_j < 0$, $j = \overline{1,n}$.

Замечание. В первом случае мы производим оценку на выброс наибольшего значения исследуемой совокупности (выборки), а во втором – наименьшего значения выборки.

В соответствии с вышеизложенной теоретической постановкой был проведен анализ «на выброс» экономических показателей деятельности предприятий, работающих на рынке рекламных услуг Самарской области за 2002, 2003 годы. В результате исследования было выявлено, что большинство показателей экономической деятельности предприятий имеет эмпирический закон распределения, близкий к теоретическому экспоненциальному распределению. Это такие показатели, как «Средняя численность работающих»; «Занятые рекламной деятельностью»; «Выручка от продажи товаров»; «Выручка от рекламной деятельности»; «Расходы, понесенные организацией»; «Расходы на рекламную деятельность»; «Материальные расходы»; «Расходы на оплату труда»; «Прочие расходы»; «Арендные платежи за арендуемые основные средства» (2002 год); «Комиссионные сборы и прочие расходы за выполненные сторонними организациями работы» (2002 год).

Число предприятий, представивших данные по каждому из показателей, различно. По показателю « Расходы на оплату труда» (2002 год) конечное число наблюдений составило 112, «Выручка от рекламной деятельности» (2003 год) и «Расходы на рекламную деятельность» (2003 год) – 103, «Материальные расходы» (2003 год) – 80, «Арендные платежи за арендуемые основные средства» (2002 год) – 25 наблюдений.

В результате визуального анализа были обнаружены выбросы в совокупностях значений таких показателей, как: «Расходы на оплату труда» (2002 год); «Выручка от рекламной деятельности» (2003 год); «Расходы на рекламную деятельность» (2003 год); «Материальные расходы» (2003 год), «Арендные платежи за арендуемые основные средства» (2002 год).

Из таблиц частот каждого показателя выявлено аномальное значение, претендующее на выброс. Это значение в несколько раз превышает среднее значение данного показателя в целом по отрасли (см. таблицу 2). С экономической точки зрения целесообразно исключить данное наблюдение из исследуемой совокупности в силу того, что оно сильно исказит результаты статистического анализа.

Этот экономический вывод мы подтвердим со статистико-математической точки зрения, а именно применением критерия Граббса для совокупностей, подчиняющихся экспоненциальному закону распределения.

Мы имеем наблюдаемое значение критерия Граббса $-G_n$. При определенном уровне значимости $\alpha = 0{,}01$, $\alpha = 0{,}05$, $\alpha = 0{,}1$ и конечном числе наблюдений n по составленной нами в соответствии с формулами (1) – (5) таблице критических значений для критерия Граббса при экспоненциальном законе распределения исследуемой совокупности (см. таблицу 1) определяем $G_{кр}$.

Проверяемое нами признаковое значение относится к классу выбросов, если $G_n > G_{кр}$, где $G_{кр} = G_{\alpha,n}$ (см. таблицу 2).

Интересен тот факт, что при проверке выделяющихся наблюдений на аномальность только часть из них действительно относится к классу «выбросов». Остальные, казалось бы, аномальные значения по данным робастного оценивания явились неотъемлемой частью исследуемых совокупностей. Это иллюстрирует анализ следующих показателей: «Средняя

Таблица 2. Применение критерия Граббса для показателей, подчиняющихся закону распределения Лапласа

Аномальное значение, тыс. руб.		Среднее значение, тыс. руб.		Конечное число наблюдений, n		$G_{кр}$		G_n		Вывод	
2002 г.	2003г.	2002 г.	2003 г.	2002 г.	2003 г.	2002 г.	2003 г.	2002 г.	2003 г.	2002 г.	2003 г.
-2156	-9137	-263,049	-710,091	41	22	4,2365 б=0,05	3,7966 б=0,05	4,3677	4,3597	Выброс	Выброс
5916	12123	368,9118	565,7468	68	79	5,7464 б=0,01	5,8524 б=0,01	5,9955	6,7494	Выброс	Выброс
-2156	-646	-193,724	-159,294	29	17	3,9918 б=0,05	3,1646 б=0,05	4,6229	2,7265	Выброс	Не выброс
5409	12123	406,1711	660,2073	76	82	4,6726 б=0,05	5,8787 б=0,01	5,3657	6,418	Выброс	Выброс

Таблица 3. Таблица критических значений критерия Граббса при законе распределения Лапласа исследуемой совокупности

Конечное число наблюдений, n	$G_{кр}$ при $\alpha = 0{,}01$	$G_{кр}$ при $\alpha = 0{,}05$	$G_{кр}$ при $\alpha = 0{,}1$
1	2	3	4
1	2,7662	1,6282	1,1380
2	3,2546	2,1093	1,6098
3	3,5407	2,393	1,8903
4	3,7438	2,5949	2,0907
5	3,9014	2,7518	2,2466
6	4,0302	2,8801	2,3743
7	4,1391	2,9887	2,4824
8	4,2335	3,0828	2,5762
9	4,3167	3,1658	2,6589
10	4,3912	3,2401	2,7330
11	4,4586	3,3073	2,8001
12	4,5201	3,3687	2,8613
13	4,5766	3,4252	2,9177
14	4,6290	3,4775	2,9699
15	4,6778	3,5262	3,0185
16	4,7234	3,5718	3,0640
17	4,7663	3,6146	3,1067
18	4,8067	3,6549	3,1470
19	4,8449	3,6931	3,1851
20	4,8811	3,7293	3,2213
21	4,9156	3,7638	3,2557
22	4,9485	3,7966	3,2885
23	4,9800	3,828	3,3199
24	5,0100	3,8581	3,3499
25	5,0389	3,8869	3,3787
26	5,0666	3,9146	3,4064
27	5,0933	3,9413	3,4330
28	5,1190	3,967	3,4587
29	5,1438	3,9918	3,4834
30	5,1678	4,0157	3,5074
31	5,1910	4,0389	3,5305
32	5,2134	4,0613	3,5529
33	5,2352	4,0831	3,5747
34	5,2563	4,1042	3,5957
35	5,2768	4,1246	3,6162
36	5,2967	4,1446	3,6361
37	5,3161	4,1639	3,6554
38	5,3349	4,1828	3,6743
39	5,3533	4,2011	3,6926
40	5,3712	4,2190	3,7105
41	5,3886	4,2365	3,7279
42	5,4057	4,2535	3,7449
43	5,4223	4,2701	3,7616
44	5,4386	4,2864	3,7778
45	5,4545	4,3022	3,7937
46	5,4700	4,3178	3,8092
47	5,4852	4,3330	3,8244
48	5,5001	4,3478	3,8392
49	5,5147	4,3624	3,8538
50	5,5290	4,3767	3,8681
51	5,5430	4,3907	3,8821
52	5,5567	4,4044	3,8958
53	5,5702	4,4179	3,9092
54	5,5834	4,4311	3,9224
1	2	3	4
55	5,5963	4,4441	3,9354
56	5,6091	4,4568	3,9481
57	5,6216	4,4693	3,9606
58	5,6339	4,4816	3,9729
59	5,6460	4,4937	3,9850
60	5,6579	4,5056	3,9969
61	5,6696	4,5172	4,0086
62	5,6811	4,5287	4,0200
63	5,6924	4,5400	4,0314
64	5,7035	4,5512	4,0425
65	5,7145	4,5621	4,0534
66	5,7253	4,5729	4,0642
67	5,7359	4,5836	4,0748
68	5,7464	4,5940	4,0853
69	5,7567	4,6043	4,0956
70	5,7669	4,6145	4,1058
71	5,7769	4,6245	4,1158
72	5,7868	4,6344	4,1257
73	5,7965	4,6442	4,1354
74	5,8061	4,6538	4,1451
75	5,8156	4,6633	4,1545
76	5,8250	4,6726	4,1639
77	5,8342	4,6819	4,1731
78	5,8434	4,6910	4,1823
79	5,8524	4,7000	4,1913
80	5,8613	4,7089	4,2001
81	5,8701	4,7177	4,2089
82	5,8787	4,7264	4,2176
83	5,8873	4,7349	4,2262
84	5,8958	4,7434	4,2346
85	5,9041	4,7518	4,2430
86	5,9124	4,7600	4,2513
87	5,9206	4,7682	4,2594
88	5,9287	4,7763	4,2675
89	5,9367	4,7843	4,2755
90	5,9446	4,7922	4,2834
91	5,9524	4,8000	4,2912
92	5,9601	4,8077	4,2989
93	5,9677	4,8153	4,3066
94	5,9753	4,8229	4,3141
95	5,9828	4,8304	4,3216
96	5,9902	4,8378	4,3290
97	5,9975	4,8451	4,3363
98	6,0048	4,8524	4,3436
99	6,0119	4,8595	4,3507
100	6,0191	4,8666	4,3578
101	6,0261	4,8737	4,3649
102	6,0331	4,8806	4,3718
103	6,0400	4,8875	4,3787
104	6,0468	4,8944	4,3856
105	6,0535	4,9011	4,3923
106	6,0603	4,9078	4,3990
107	6,0669	4,9145	4,4057
108	6,0735	4,9211	4,4122
109	6,0800	4,9276	4,4187
110	6,0864	4,9340	4,4252
111	6,0928	4,9404	4,4316
112	6,0992	4,9468	4,4379

численность работающих» (2002, 2003 гг.), «Занятые рекламной деятельностью» (2002, 2003 гг.), «Выручка от продажи товаров» (2002, 2003 гг.), «Выручка от рекламной деятельности» (2002 г.), «Расходы, понесенные организацией» (2002, 2003 гг.), «Расходы на рекламную деятельность» (2002 г.), «Материальные расходы» (2002 г.), «Расходы на оплату труда» (2003 г.), «Комиссионные сборы и прочие расходы за выполненные сторонними организациями работы» (2002 г.).

При анализе данных показателей наблюдаемые значения критерия Граббса оказались меньше критического значения, а именно $G_{кр} > G_n$, где $G_{кр} = G_{\alpha.n}$. Следовательно, исследуемые аномальные наблюдения не могут быть отнесены к классу выбросов и являются неотъемлемой частью изучаемой нами совокупности.

В процессе данного исследования был выявлен ряд показателей с эмпирическим законом распределения, близким к теоретическому закону распределения Лапласа. Это такие показатели, как «Прибыль» (2002, 2003 гг.), «Прибыль от рекламной деятельности» (2002, 2003 гг.).

Установлено, что в каждой совокупности значений показателей имеются максимальные и минимальные значения, которые «подозрительны» на выброс. Эти значения существенно отличаются от среднего значения анализируемого показателя в целом по отрасли. Данные аномальные наблюдения мы подвергнем анализу с использованием критерия Граббса.

Особенностью исследования данных совокупностей является то, что они подчиняются закону распределения Лапласа. Исходя из этого, мы делим совокупность значений на две части – положительную и отрицательную. В каждой из этих частей присутствует аномальное значение, которое будет нами исследовано на возможность его отбраковки. Данные анализа представлены в таблице 2.

В процессе использования критерия Граббса на выброс мы аналогично случаю с совокупностями, имеющими экспоненциальный закон распределения, сравниваем наблюдаемые значения критерия G_n с его критическим значением при определенном конечном числе наблюдения n и заданном нами уровне значимости α. При $G_n > G_{кр}$, где $G_{кр} = G_{\alpha,n}$, аномальное значение относится к классу выбросов. В противном случае оно является часть исследуемой совокупности.

Как видно из таблицы 2, большинство аномальных значений показателей прибыли (как максимальных, так и минимальных) явились выбросами, за исключением одного: минимального значения показателя «Прибыль от рекламной деятельности» за 2003 год.

Критические значения критерия Граббса для определенного конечного числа наблюдений и заданного исследователем уровня значимости представлены в построенной нами таблице критических значений данного критерия при законе распределения Лапласа исследуемой совокупности (см. таблицу 3).

Таким образом, мы показали, что применение критерия Граббса для отбраковки аномальных значений из имеющегося массива первичных статистических данных возможно и в том случае, когда исследуемая совокупность не подчиняется теоретическому нормальному закону распределения. В нашем случае это экспоненциальное распределение и распределение Лапласа. Нами была решена задача построения таблиц критических значений данного критерия для любого конечного числа наблюдений. В результате мы получили массив статистических данных, исключающий наличие выбросов и пригодный для дальнейшего глубокого статистического анализа.

Список использованной литературы:

1. Дубров А.М., Мхитарян В.С., Трошин Л.И. Многомерные методы статистики: Учебник. – М.: «Финансы и статистика», 2000. – 350 с.
2. Хьюбер Дж. Робастность в статистике. – М.: «Мир», 1984. – 304 с.
3. Прикладная статистика. Основы эконометрики: Учебник для вузов: В 2 т. 2-е изд., испр. – Т. 1.
4. Айвазян С.А., Мхитарян В.С. Теория вероятностей и прикладная статистика. – М.: Юнити – Дана, 2001. – 656 с.