

**МОСКОВСКИЙ ГОСУДАРСТВЕННЫЙ ТЕХНИЧЕСКИЙ
УНИВЕРСИТЕТ ИМЕНИ Н. Э. БАУМАНА**
Факультет информатики и систем управления
Кафедра теоретической информатики и компьютерных технологий

Лабораторная работа №4
по курсу «Математическое моделирование»

«Определение выброса методом исключения одной точки»

Выполнил:
студент ИУ9-111
Выборнов А. И.
Руководитель:
Домрачева А. Б.

Москва 2015

1. Постановка задачи

Рассматриваются 6 станций чешского метрополитена. Для каждой станции, вручную была посчитана следующая информация с точностью до месяца:

- Среднее число пассажиров, вошедших с данной станции в метрополитена в день.
- Среднее число пассажиров, вышедших с данной станции в день.

Данные для 6 станций (A0, A1, B0, B1, C0, C1) приведены в таблице 1. Строки соответствуют месяцам, столбцы станциям метро, причём префикс "th" соответствует вошедшим пассажирам, а префикс "r" вышедшим.

Таблица 1: Данные о числе пассажиров проходящих через станции Пражского метро

m/s	thA0	rA0	thA1	rA1	thB0	rB0	thB1	rB1	thC0	rC0	thC1	rC1
1	16551	14899	30746	27320	32822	29553	21002	18793	17084	15365	4544	3118
2	16810	14292	22558	20155	25314	22567	40022	35436	29096	25876	17519	16162
3	14434	13046	28001	24916	36918	32720	35118	31145	38639	34226	38841	34819
4	20891	18696	32958	29255	46677	41259	20283	18164	23690	21145	37324	33492
5	13773	12468	28277	25159	16909	15212	41746	36944	29087	25868	16717	15461
6	14739	13313	36763	32398	21889	19569	40458	35817	21993	20494	40099	35920
7	24713	22040	34650	30735	34998	31040	19478	17460	30082	26738	42244	37797
8	10127	9278	33590	29808	23285	20791	22974	21353	18776	17263	22099	20170
9	14689	13269	12239	11126	21561	19282	25348	23430	34808	31290	40895	36617
10	13047	11833	35848	31784	37778	33472	25336	22586	26192	23751	17519	16162
11	16487	14843	38451	34061	29376	26120	23743	22025	18230	16784	38841	34819
12	14345	12968	18573	16668	32822	29553	29751	27282	37085	33283	37324	33492

Для обобщённых данных необходимо сравнить функции распределения для вошедших и вышедших, основываясь на критерии Колмогорова-Смирнова. Определить выброс, методом исключения одной точки.

2. Решение

Область значений входных данных была отображена на интервал от 0 до 1. Область определения входных была разбита на 20 интервалов и, на основе этого, были посчитаны функции распределения. Для построенных функций распределения использовалась модификация критерия Колмогорова-Смирнова — вместо функции $\max_i |x_i - y_i|$ используется функция $\sum_i |x_i - y_i|$. Это делается с целью получить интегральную характеристику расхождения функций распределения, что позволяет в рамках задачи получить более чувствительный критерий.

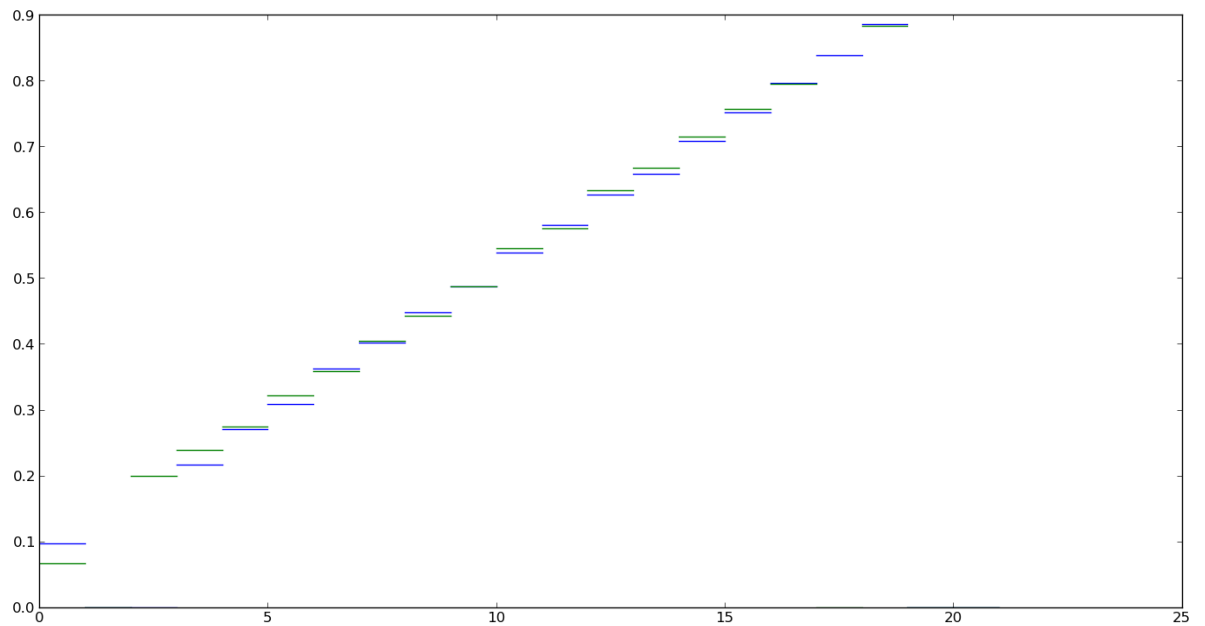


Рисунок 1 — Функции распределения: синий график вошедшие пассажиры, зелёный график вышедшие

Полученные функции распределения изображены на рисунке 1. Согласно модифицированному критерию полученные функции распределения расходятся на значение: 1.16281626925.

Попытаемся найти выброс методом исключения одной точки. Для этого посчитаем меру расхождения функций согласно модифицированному критерию Колмогорова-Смирнова для входных данных, последовательно исключая из них по одной точке. Результаты, отсортированные по мере расхождения, приведены в таблице 2.

Из полученных метрик минимальное значение 1.0832150432 достигается в точке (4544.0, 3118.0). Исключая из входных данных точку (4544.0, 3118.0), мы получили функции распределения, изображённые на рисунке 2.

Исключением точки (4544.0, 3118.0) мы не только снизили значение коэффициента расхождения функций распределений, но и, как видно из графика, получили на порядок более однородную картину. Также если рассмотреть минимальные коэффициенты расхождения из таблицы видно, что второе по минимальности значение меры расхождения уже на порядок ближе к значению функции расхождения для исходных данных. Поэтому полагаем, что это точка, является точкой выброса.

Таблица 2: Поиск выброса методом исключения одной точки

кол-во вошедших для исключённой точки	кол-во вышедших для исключённой точки	мера расхождения
17519	16162	1.0832150432
42244	37797	1.15575747513
37778	33472	1.15629878384
38841	34819	1.15855455986
38841	34819	1.15855455986
32822	29553	1.15863648049
35118	31145	1.15867968521
38639	34226	1.15902746535
21993	20494	1.15904603266
10127	9278	1.16050892302
36763	32398	1.16063853718
26192	23750	1.1608474195
29096	25876	1.16089210348
37085	33283	1.16092704473
40458	35817	1.16093632838
14738	13313	1.16097132064
32822	29553	1.1610998636
21002	18793	1.16112557219
25314	22567	1.16112557219
28001	24916	1.16122277006
40099	35920	1.16160398667
29376	26119	1.16165754624
30746	27320	1.16189254522
16717	15461	1.16197625118
20891	18696	1.16203501368
21889	19569	1.1620832683
13047	11833	1.16244527997
23690	21145	1.16245135006
24713	22040	1.16252454814
40895	36617	1.16269183252
21561	19282	1.162862407
40022	35436	1.16290981997
33590	29808	1.16297373439
36918	32720	1.16307728289
23285	20791	1.16320404053
34998	31040	1.16326777642
20283	18164	1.16328437989
23743	22025	1.1633158015
4544	3118	1.16363996442
18573	16668	1.1637618762
38451	34061	1.16377177196
13773	12468	1.16422202941
28276	25158	1.16422488592
30082	26737	1.16440451451
16810	14292	1.16452007566
12239	11126	1.16478904671
14345	12967	1.16495589752
25347	23430	1.16508273168
34808	31290	1.16509897808
18776	17263	1.16535356457
29086	25868	1.16561873544
46677	41259	1.1660786081
14434	13046	1.1663282467
19478	17460	1.16664863495
29751	27282	1.16684285214
25336	22586	1.16701783891
37324	33492	1.16704461869
37324	33492	1.16704461869
34650	30735	1.16772824283
17084	15365	1.16782016125
17519	16162	1.16794549064
22974	21353	1.17095625254
41746	36944	1.1714132942
22099	20170	1.17327288243
16487	14843	1.17568413423
32958	29255	1.1761222515
14689	13269	1.18100509887
18230	16784	1.18655529847
35848	31784	1.35960840814
22558	20155	1.82476255258
16909	15211	2.46879736834

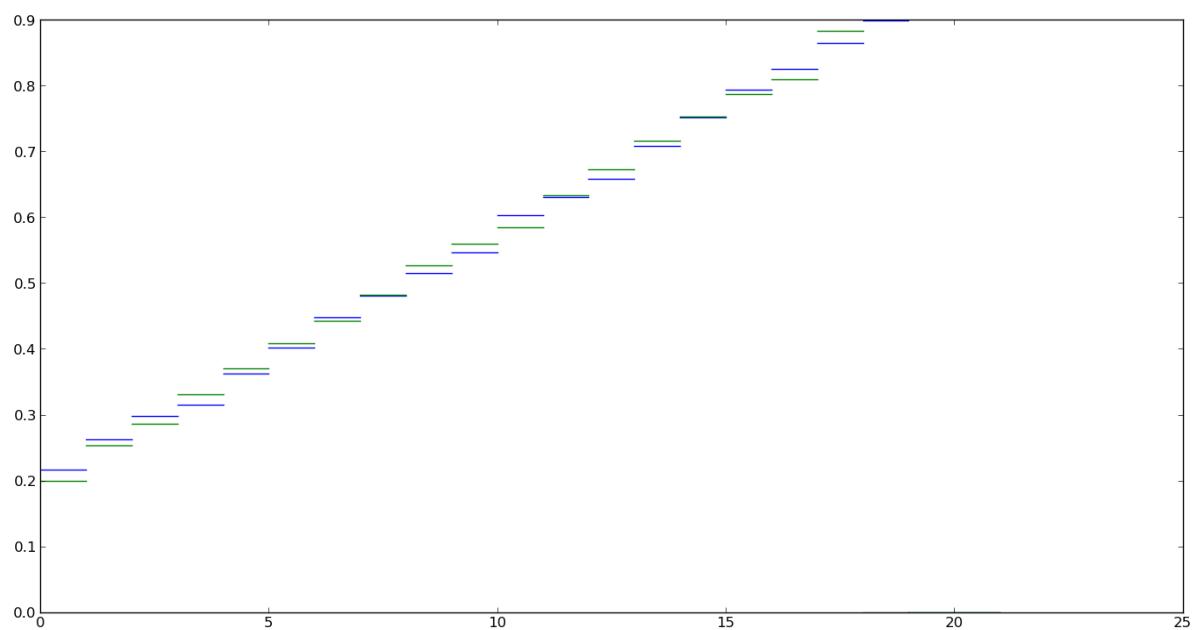


Рисунок 2 — Функции распределения: синий график вошедшие пассажиры, зелёный график вышедшие