Датасет содержит:

- refined dataset: 5316 реакций, в них участвует 1412 различных белков, 3913 малых молекул;
- general dataset: 19443 реакции;
- для каждой реакции указан участвующий белок, участвующая малая молекула, ссылка на pdf-референс, разрешение, и один количественный показатель - либо константа ингибирования, либо константа диссоциации, либо IC50 (концентрация полумаксимального ингибирования).

Для оценки статистистических показателей датасеты разделялись на несколько подгрупп в зависимости от предоставленных показателей.

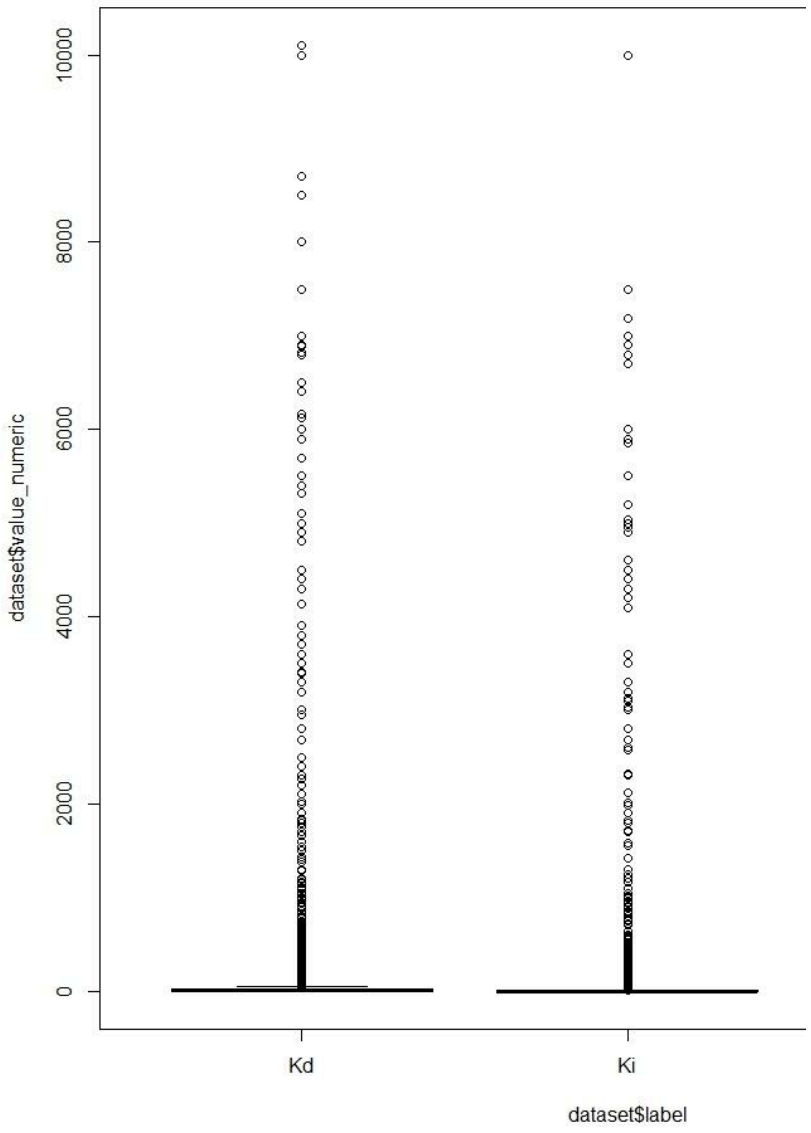
Базовые статистические показатели Refined выборки (данные в пересчёте на наномоли):

Выборка	Среднее	Медиана	Стандартное отклонение
выборка с Kd	131,65	0,11	692,62
выборка с Ki	177,91	1,00	827,04

Выборка	Среднее	Медиана	Стандартное отклонение
выборка с Kd	1248,95	1,7	5273,81
выборка с Ki	665,68	0,14	9101,08
выборка с IC50	246,64	0,16	5883,27

Ниже приведены боксплоты для датасетов, позволяющие приблизительно оценить распределение данных. Все значения приведены в наномолях.

рис.1. “ящики с усами” для refined выборки



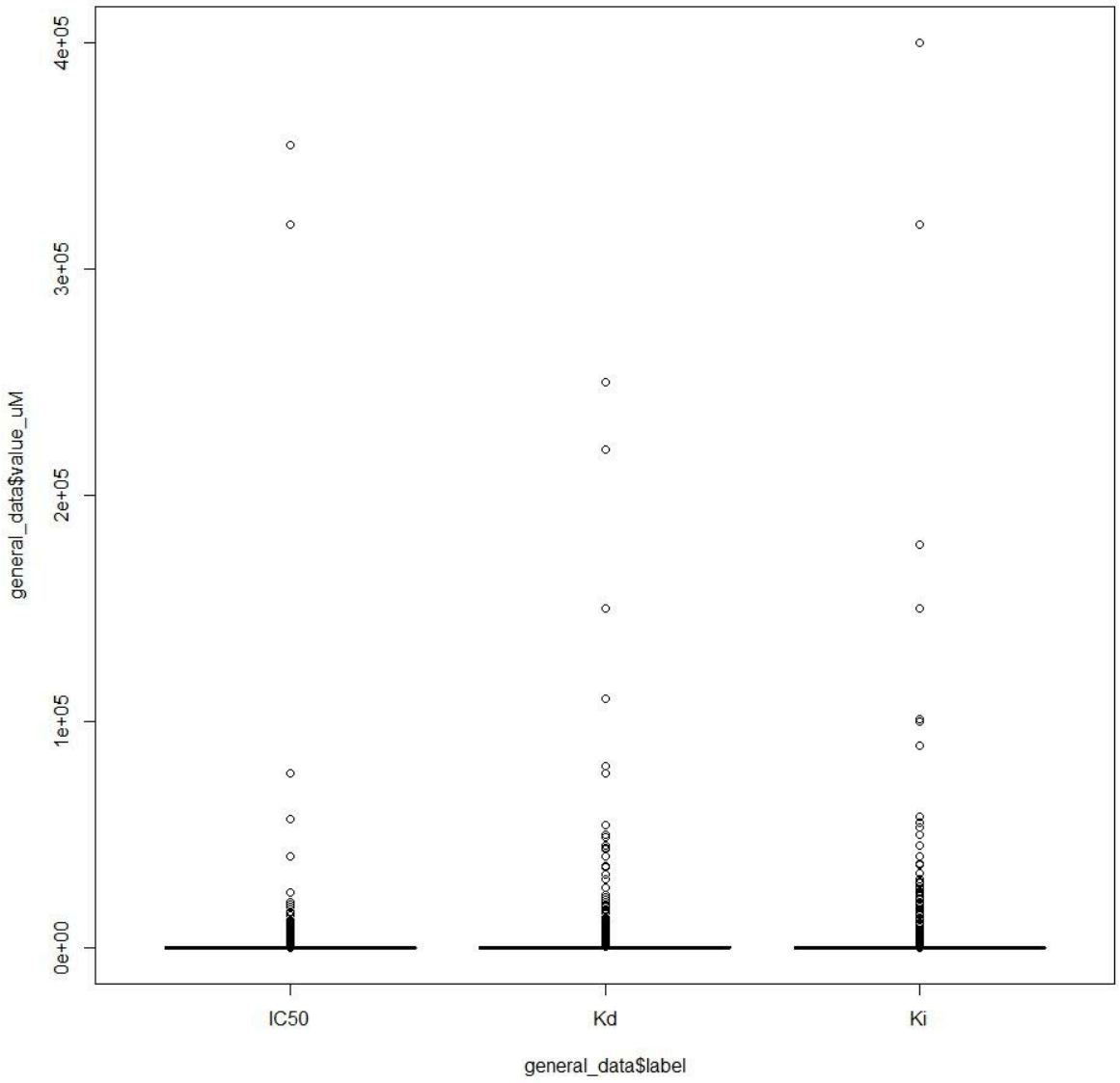


рис.2. “ящики с усами” для general выборки

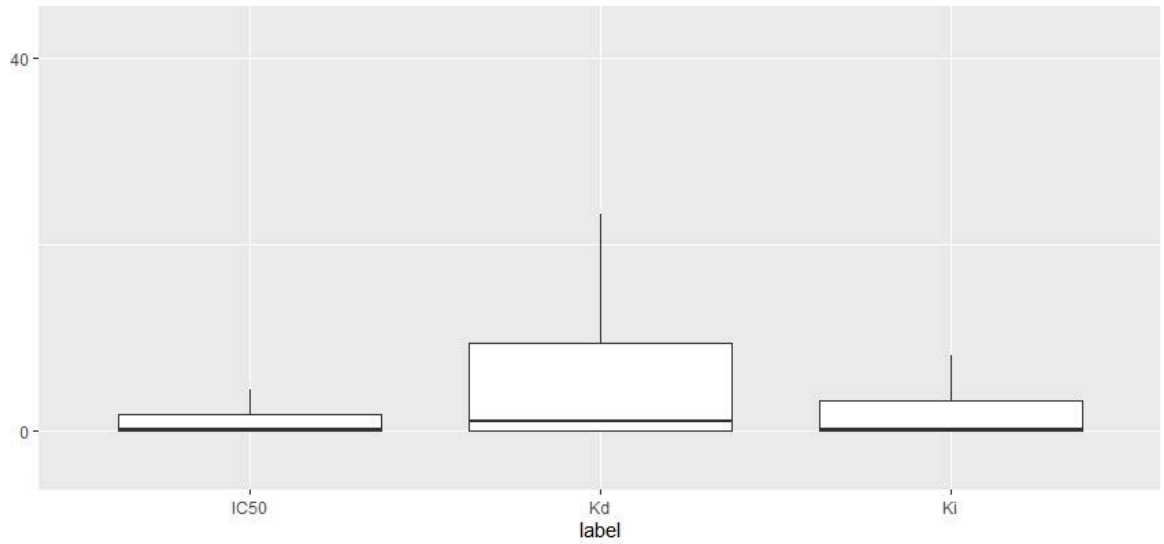


рис.3. “ящики с усами” для general выборки, без учета выпадающих значений (приблизительно 3000 значений, что составляет примерно ⅙ часть датасета)

Проверка на нормальность
(тест Шапиро-Уилка на refined выборке)

Для датасета, для которого определена константа ингибирования:

W = 0.18682, p-value < 2.2e-16

Для датасета, для которого определена константа диссоциации:

W = 0.21695, p-value < 2.2e-16

p-value очень низкое, данные распределены ненормально, что подтверждается Q-Q графиками, на которых данные очень удалены от теоретического распределения:

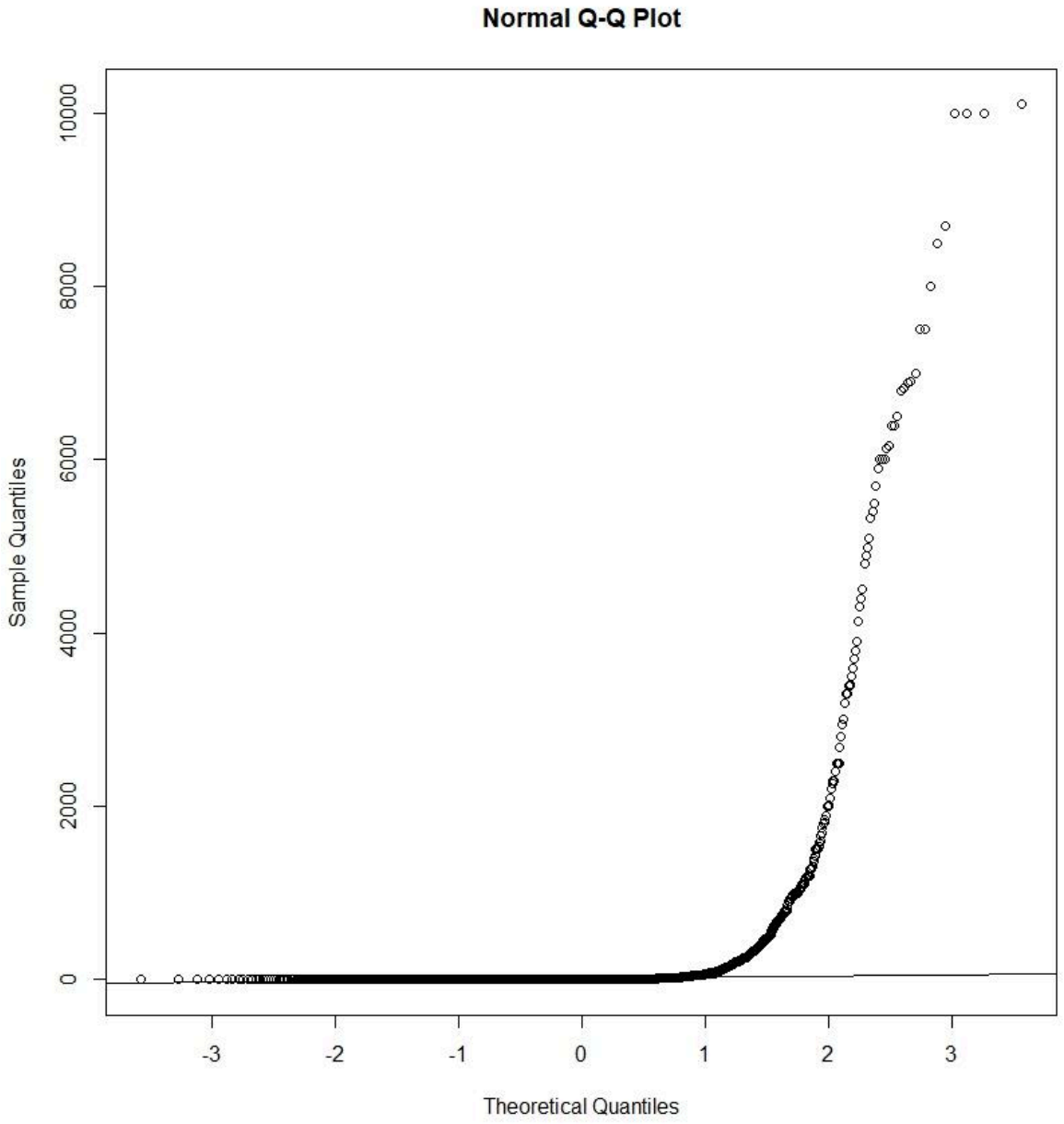


рис.4. QQ-plot для refined Ki выборки

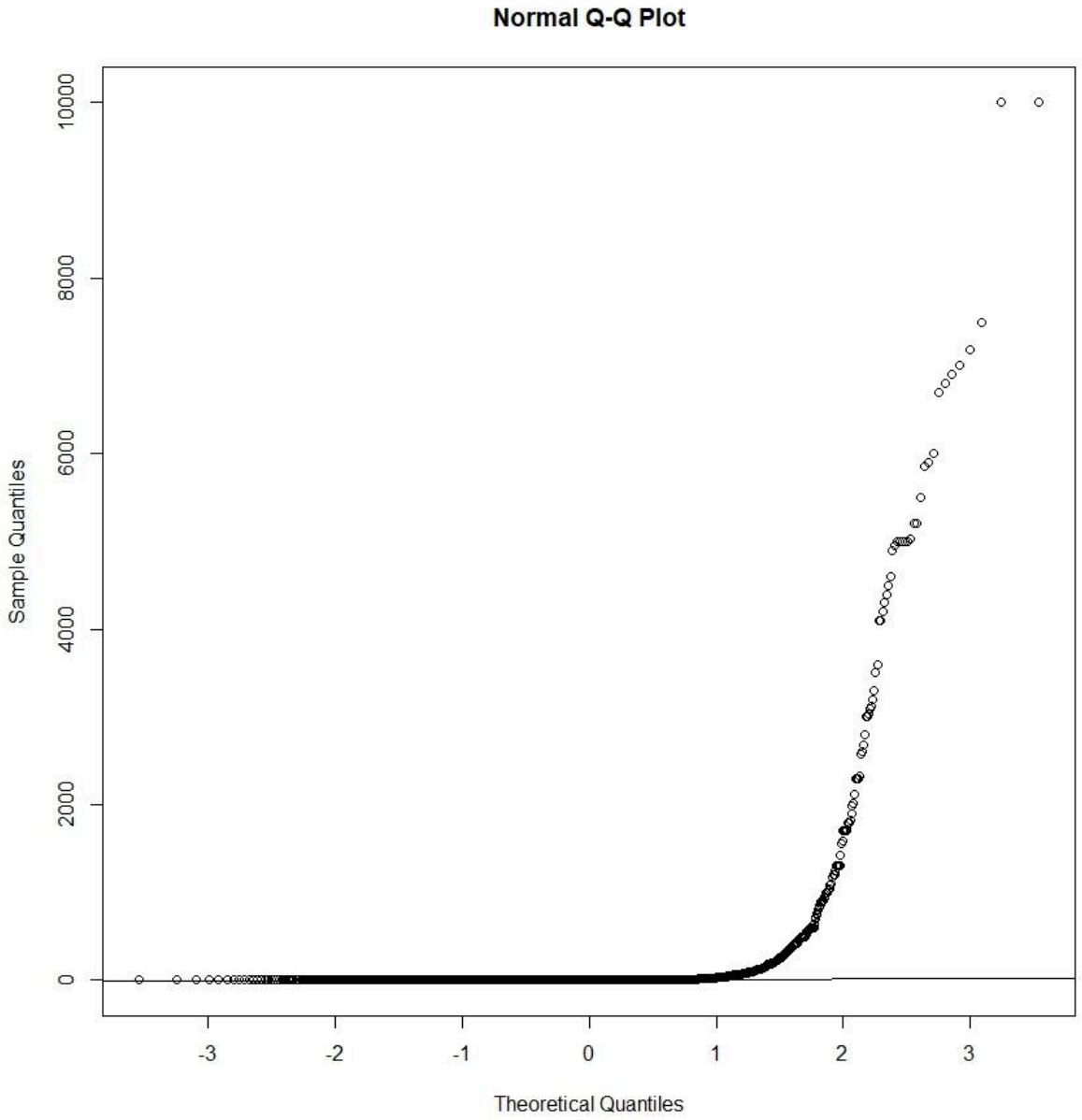


рис.5. QQ-plot для refined Kd выборки

General:

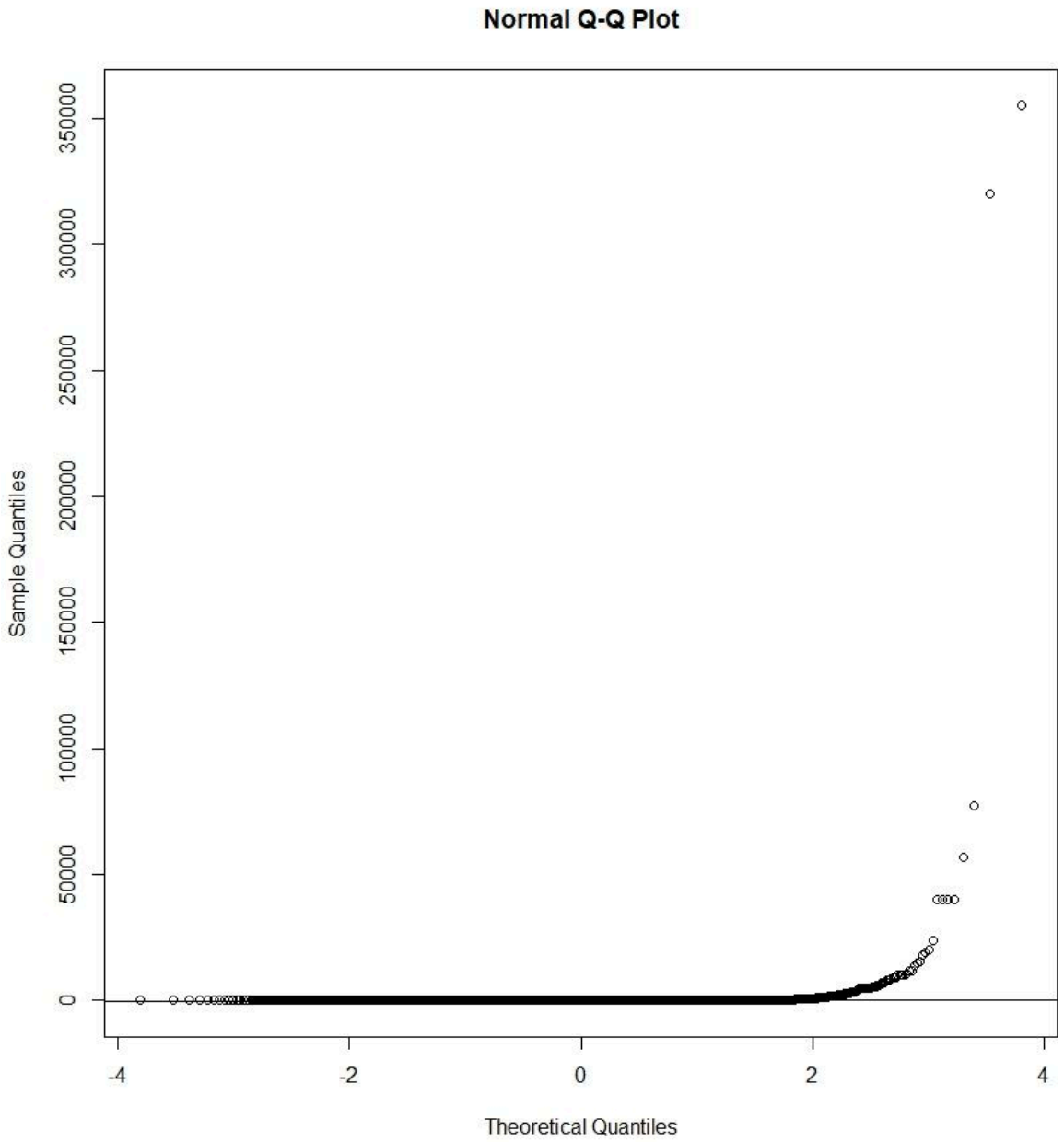


рис.6. QQ-plot для general IC50 выборкирис.6. QQ-plot для general IC50 выборкирис.6. QQ-plot для general IC50 выборкирис.6. QQ-plot для general

IC50 выборкирис.6. QQ-plot для general IC50 выборкидfsd

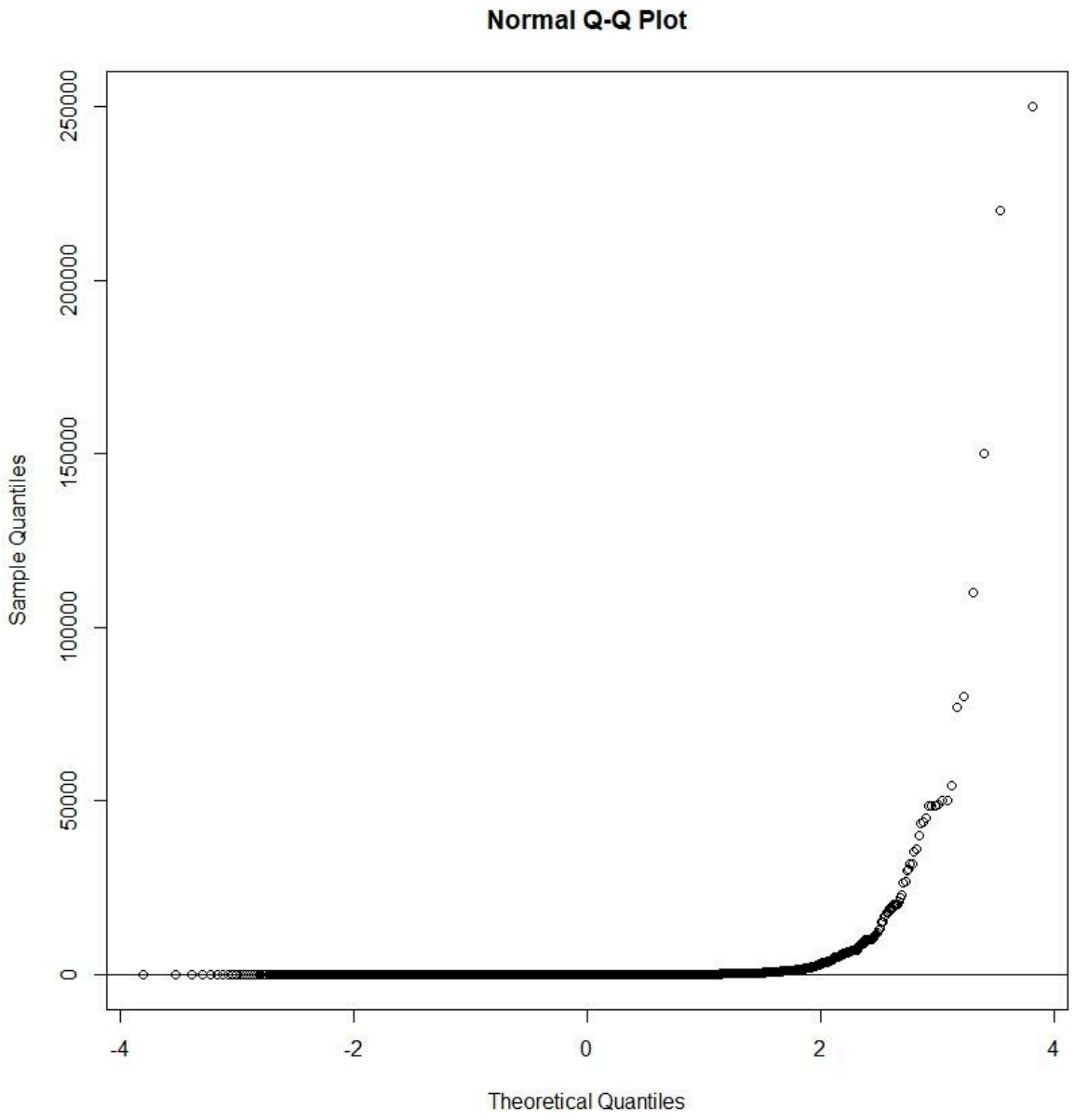


рис.7. QQ-plot для general Kd выборки

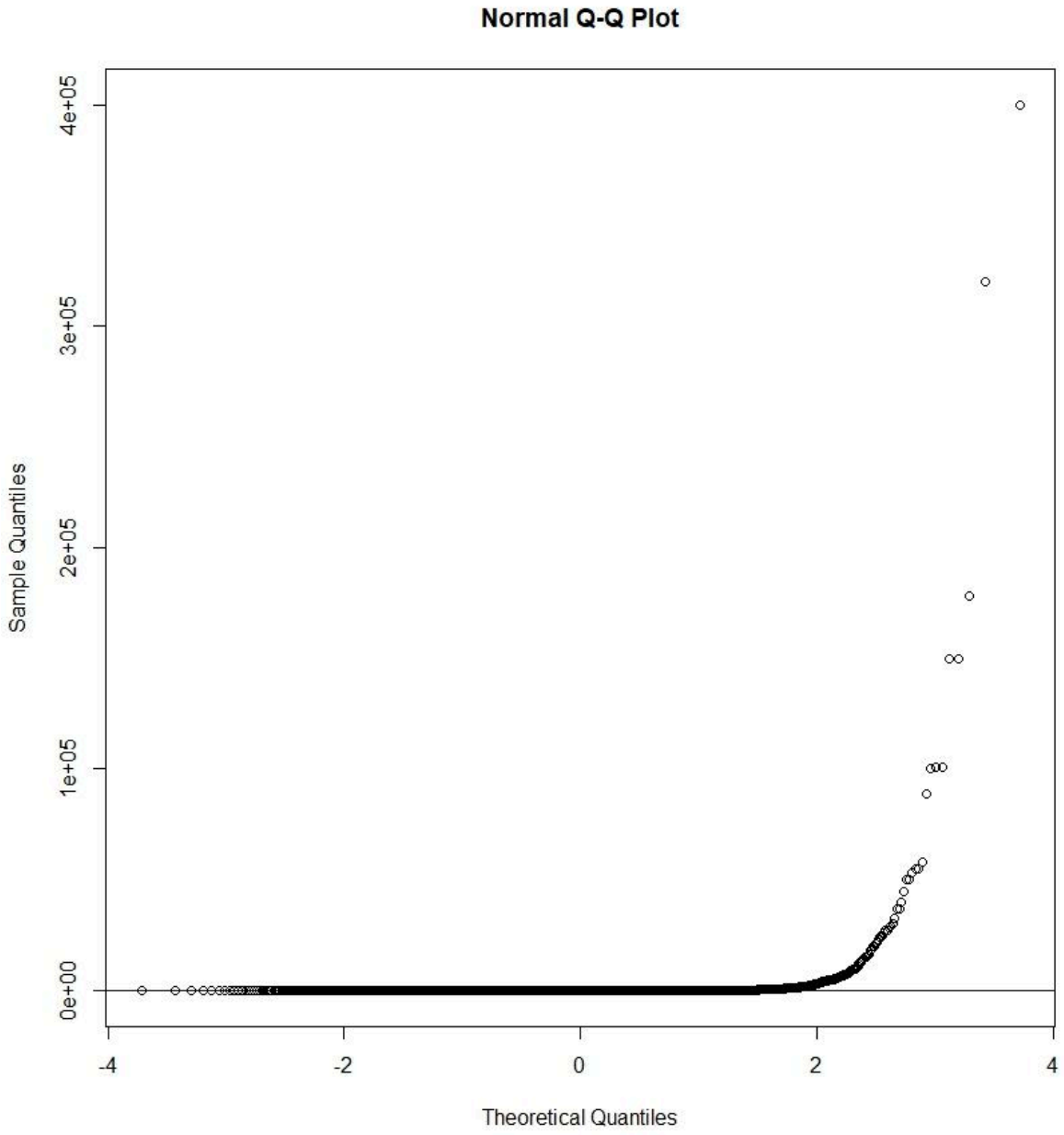


рис.8. QQ-plot для general Ki выборки