1. Оценка значимости уравнения регрессии
   1. Оценка значимости коэффициентов регрессии

При проверке гипотез пользуются принципом практической уверенности. Критическое значение статистики критерия – такое значение, что если гипотеза верна, то вероятность мала. Иначе говоря, вероятность отвергнуть нулевую гипотезу достаточно мала в условиях данного исследования.

Значимость коэффициентов регрессии проверяется следующим образом.

Считается t-статистика. Если , то отвергается (коэффициент значим) на уровне значимости . Таким образом, вероятность совершить ошибку первого рода, неверно отвергнув гипотезу (приняв коэффициент значимым, в то время как на самом деле он незначим, близок к нулю) равна .

1. Кластеризация по структуре
   1. Низкое качество модели на кластере №3

*Возможная причина* – нарушение однородности объектов в кластере. В частности, точки, принадлежащие отрасли G находятся слишком далеко по всем трем переменным модели (рис. 1, оранжевое)

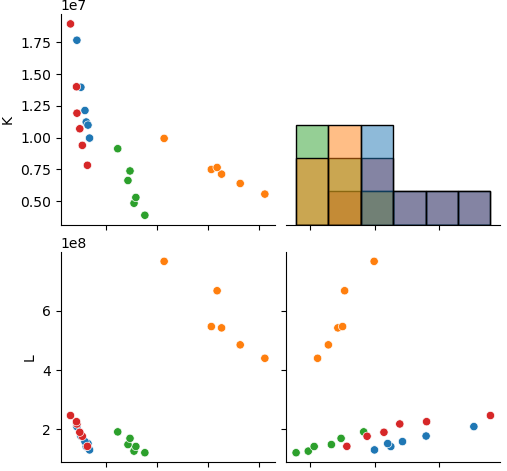


Рис 1 – Отрасль G является выбросом в кластере 3

*Возможное решение* – исключить отрасль G из рассмотрения. Работает. Тогда ключевым является разделение на группы, однородные по исходным признакам, а не по любым другим. В таком случае почему кластеры, построенные по исходным признакам неравномерны по размеру? На рисунке 2 видно, что синий кластер содержит намного больше отраслей.

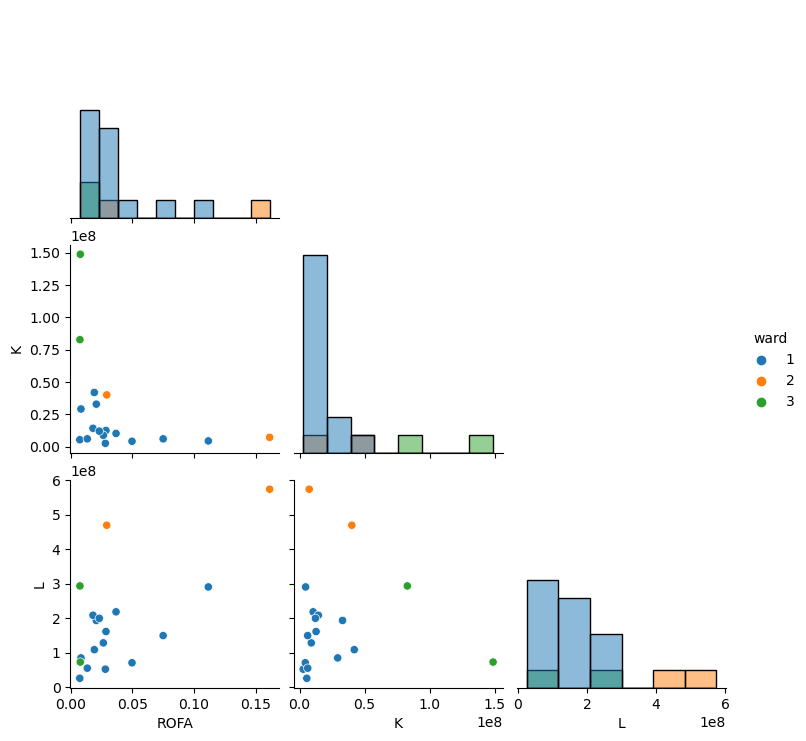


Рис. 2 – плохая кластеризация

Обосновать исключение отрасли G. Посмотреть на профили кластеров и убедиться, что отрасль G – выброс? Можно сделать это графически/с помощью расстояния Махаланобиса. Априорное или апостериорное определение аномалий?

TODO

Описать таблицы!

Сделать адекватную кластеризацию по исходным признакам. Что значит «адекватная»?

Проблема репрезентативности кластеризации в задаче регрессии.

1. Как связана -статистика с характеристиками выборки? Чем больше разброс, тем меньше t-статистика и тем меньше вероятность того, что коэффициент окажется значимым