

Дз-2 EmpIO (Сысоев Никита)

Задание 1

Для определения количества картелей нужно определить, находится ли определенный набор компаний в картельном сговоре.

Проблемы обнаружении картелей:

- иногда сложно определить длительность картельного соглашения (особенно его окончание)
- нужно учитывать влияние внешних факторов на цены
- время причинения вреда и совершения нарушения могут не совпадать
- для ряда методов требуется найти сопоставимый рынок, на котором нет картеля.

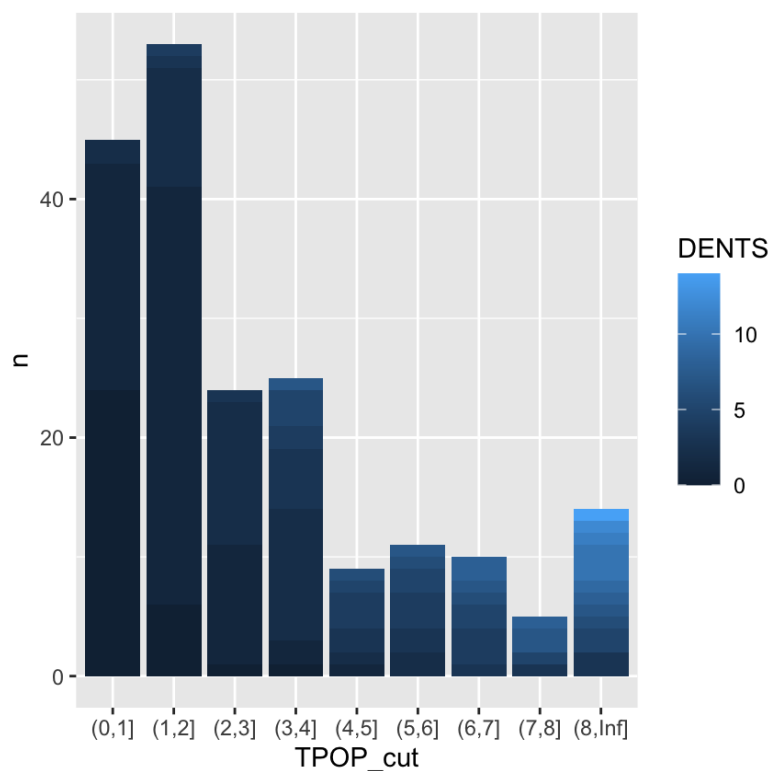
Проблемы с определением вероятности обнаружения картеля:

- мы не знаем о нераскрытых картелях (в противном случае они были бы раскрыты).

Для определения вероятности раскрытия картеля можно использовать подход, описанный в статье ["A Tip of the Iceberg? The Probability of Catching Cartels"](#).

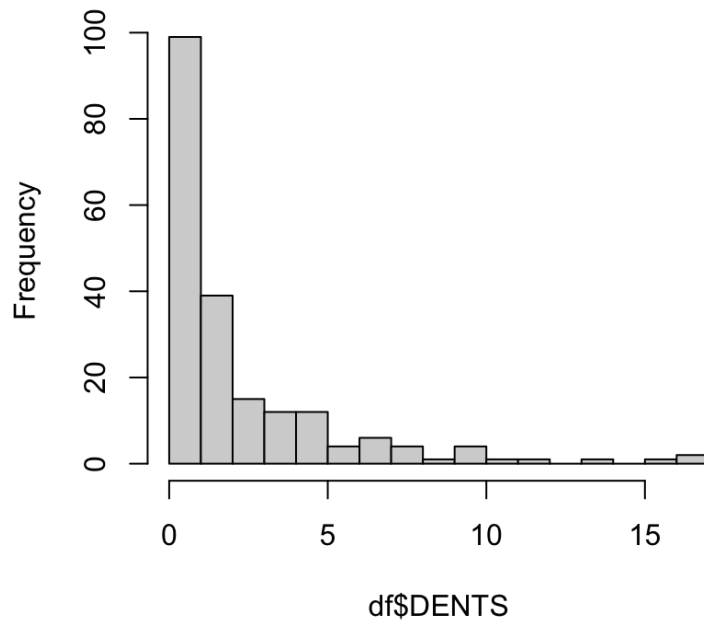
Задание 2

Пункт 1



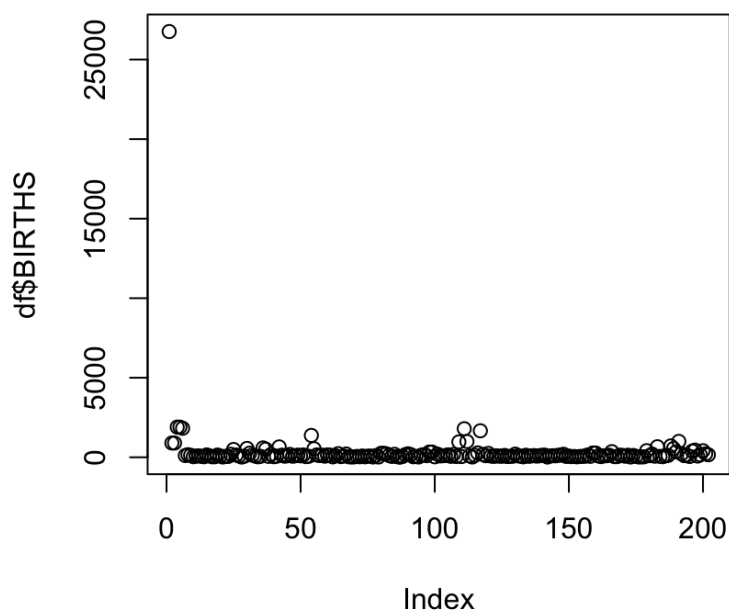
На данном графике показано распределение городов (ось ординат) по населению (ось абсцисс), а цвет характеризует количество стоматологов в соответствующих городах^[1].

Histogram of df\$DENTS



Гистограмма распределения количества городов по количеству стоматологов в них (почему-то г склеивает количество городов в которых 0 и 1 стоматологов). Городов, в которых нет стоматологов, 32, а городов, в которых 1 стоматолог, 67. Таким образом, **наиболее часто встречаемое число стоматологов в одном городе = 1. Медианное число стоматологов в одном городе = 2** (для данных без выбросов медиана = 1).

В результате устранения выбросов были убраны 6 наблюдений (количество сократилось с 202 до 196). Стоит отметить, что очень сильный выброс наблюдался в данных по рождениям (см. график), поэтому его точно надо устранить для получения разумных результатов регрессии и логит-модели.



Пункт 2

Таблица с корреляциями

	DENTS	DENTS	Интерпретация	Гипотеза
TPOP	0.86	0.86	Высокая корреляция с населением	Чем больше людей, тем больше пациентов у стоматолога
BIRTHS	0.16	0.53	Значительная корреляция между количеством стоматологов и рожденьями	Возможно, стоматологи предпочитают оседать в городах с большой рождаемостью, чтобы быть уверенными в потоке клиентов в долгосрочной перспективе
ELD	-0.26	-0.28		Возможно, пожилые люди не целевая аудитория стоматологов, потому что у них низкая готовность платить
<i>outliers removed</i>	<i>no</i>	<i>yes</i>		

Пункт 3

	DENTS	DENTS
(Intercept)	−8.758***	−10.370***
	(2.157)	(1.803)
TPOP	0.367***	0.520***
	(0.041)	(0.047)
OPOP	0.398*	−0.438
	(0.239)	(0.317)
NGRW	−0.154	−0.273
	(0.724)	(0.591)
PGRW	0.224	0.272
	(0.198)	(0.164)
OCTY	1.553***	1.939***
	(0.467)	(0.415)
BIRTHS	−0.0005***	0.0001
	(0.0001)	(0.001)
ELD	−2.296	−0.413
	(2.253)	(1.899)
PINC	0.402***	0.237***
	(0.093)	(0.077)
LNHDD	0.839***	1.042***
	(0.233)	(0.190)
LANDV	0.397	0.924**
	(0.489)	(0.401)
<i>outliers removed</i>	<i>no</i>	<i>yes</i>

Будем интерпретировать коэффициенты, которые значимы на 5%-м уровне значимости (см. *столбец 2*):

- численность населения города при прочих равных увеличивает количество стоматологов (0.520)
 - чем больше людей, тем больше пациентов у стоматолога
- число лиц, работающих за пределами округа, при прочих равных увеличивает количество стоматологов (1.939)
 - возможно, многие люди, работающие на прилегающих к городу территориях, наблюдаются у стоматологов города, что увеличивает спрос на услуги стоматолога ⇒ количество стоматологов

- увеличение подушевого дохода стимулирует рост числа стоматологов при прочих равных (0.237)
 - рост доходов повышает готовность людей платить за дорогостоящие услуги стоматологов
- градусо-дни отопительного периода (чем холоднее снаружи, тем показатель больше) при прочих равных увеличивают количество стоматологов (1.042)
 - возможно, стоматологи ездят на заработки в северные регионы
 - сомнительно, но окей
- стоимость акра с/х земель и зданий при прочих равных увеличивает количество стоматологов (0.924)
 - сложно интерпретировать (возможно, в данных остались выбросы).

При этом переменные численность населения в пределах 10 миль от города, отрицательное / положительное изменение численности населения, рождаемость на 1000 жителей, доля населения старше 65 лет от общей численности населения оказались незначимыми, что в целом логично, т.к. не влияет напрямую на издержки профессиональной деятельности стоматологов или спрос на их услуги (возможно, эффект рождаемости на количество стоматологов все же есть, но он долгосрочный, и из-за этого нужно включать далекие лаги переменной).

Итак, факторы, влияющие на количество стоматологов в городе:

- численность населения города и число людей, работающих за пределами округа
- подушевой доход
- градусо-дни отопительного сезона (сложно придумать убедительную версию)
- стоимость земли и зданий (сложно придумать убедительную версию).

Почему линейная модель хуже, чем `ordered logit` / `probit`, см. *Пункт 5*.

Пункт 4

	DENTS_F	DENTS_F	DENTS_F	DENTS_F
TRDP	1.403***	1.430***	0.739***	0.520***
	(0.144)	(0.160)	(0.081)	(0.047)
ORDP	−0.516	−0.959	−0.450	−0.438
	(0.403)	(0.636)	(0.358)	(0.317)
NGRW	1.091	0.877	0.525	−0.273
	(1.039)	(1.087)	(0.624)	(0.591)
PGRW	−0.588	−0.582	−0.340	0.272
	(0.381)	(0.406)	(0.242)	(0.164)
OCTY	3.607***	3.412***	1.917***	1.939***
	(1.121)	(1.263)	(0.733)	(0.415)
BIRTHS	−0.001	−0.00004	0.0001	0.0001
	(0.0003)	(0.001)	(0.001)	(0.001)
ELD	−2.576***	−2.006	−0.632	−0.413
	(0.084)	(2.604)	(2.031)	(1.899)
PINC	0.447***	0.448***	0.246***	0.237***
	(0.145)	(0.148)	(0.085)	(0.077)
LNHDD	1.933***	1.986***	1.128***	1.042***
	(0.095)	(0.248)	(0.216)	(0.190)
LANDV	2.745***	2.933***	1.504***	0.924**
	(0.859)	(0.933)	(0.443)	(0.401)
<i>kind</i>	<i>Logit</i>	<i>Logit</i>	<i>Probit</i>	<i>LinReg</i>

	DENTS_F	DENTS_F	DENTS_F	DENTS_F
<i>outliers removed</i>	<i>no</i>	<i>yes</i>	<i>yes</i>	<i>yes</i>

На 5%-м уровне значимости наборы значимых переменных в линейной регрессии порядковом логите (и пробите) одинаковы (см. *столбцы 2, 3 и 4*). Все коэффициенты при значимых переменных увеличились по сравнению с линейной регрессией.

Пункт 5

Для оценки входа линейная модели подходит хуже порядкового логита / пробита, потому что:

- линейная модель не имеет ограничений крайних значений (0 и 1)
- по мере усиления конкуренции пороги возрастают, поэтому вероятность нелинейным образом зависит от включенных в модель факторов.

1. Данный график приведен для данных с устраненными выбросами, но график с выбросами выглядит похожим образом ↩