lab3

2024-05-30

DUMMY VARIABLES ДЛЯ КАТЕГОРІАЛЬНИХ ЗМІННИХ

```
hotel_meal <- hotel %>% mutate(M1 = (type_of_meal_plan == "Meal Plan 1"), M2 = (type_of_meal_plan == "Meal Plan 2
"), M3 = (type_of_meal_plan == "Meal Plan 3"))
```

```
hotel_room <- hotel %>% mutate(R2 = (room_type_reserved == "Room_Type 2"), R3 = (room_type_reserved == "Room_Type
3"), R4 = (room_type_reserved == "Room_Type 4"), R5 = (room_type_reserved == "Room_Type 5"), R6 = (room_type_rese
rved == "Room_Type 6"), R7 = (room_type_reserved == "Room_Type 7"))
```

```
hotel_market <- hotel %>% mutate(Online = (market_segment_type == "Online"), Corporate = (market_segment_type ==
"Corporate"), Complementary = (market_segment_type == "Complementary"), Aviation = (market_segment_type == "Aviat
ion"))
```

ПИТАННЯ НА РЕГРЕСІЮ

- 1. Що впливає на появу особливих побажань? (з “Чи є істотною наявність особливих побажань?”)
- 2. Чим зумовлено скасування/нескасування резервації? (з “Які характерні риси скасованих записів?”)

```
view(hotel)
```

Питання №1

1. Що впливає на появу особливих побажань?

Першу модель (log) побудуємо з досить інтуїтивно очікуваними регресорами.

- В прешу чергу нам цікаво дізнатися чи враховується таки потреба у паркувальному місці як особливе побажання. Окрім цього, можна очікувати, що на наявність особливих побажань впливатиме ціна, що заплачена за кімнату.
- Достовірно невідомо що саме являють собою особливі побажання, але це може бути як ранкова корзинка фруктів під дверима так і лебідь з рушників, що очікуватиме гостей на двуспальному ліжку, тому досить важливо було б врахувати ціну, яка заплачена за заброньваний номер.При побудові моделі варто врахувати, що для avg_price_per_room для адекватнішого аналізу його впливу необхідно взяти логаритм цього регресора, адже межі змінної досить широкі (від 9 до 540), через що вплив збільшення значення ціни на 1 одиницю буде майже непомітним. Тож було б логічно розглянути вплив збільшення ціни на 1 відсоток відносно наявності особливих побажань.
- Для початкової моделі братимемо дорослих і дітей поокремо. У нашому датасеті є записи з просто дорослими (їх більшість. 23 тисячі записів мають двох дорослих без дітей), з просто дітьми, і з дорослими та дітьми одночасно. На перший погляд може здатися, що вплив виключно дорослих і виключно дітей має давати більший вплив на наявність особливих побажань, адже це записи або з “парочкою”, або з виключно дітьми, для яких можуть бути передбачені окремі умови, тому фактор взаємодії між цими двома регресорами лишимо на потім.
- Останнім регресором в початкову модель додамо бінарну змінну repeated_guest, бо є підозра що у повторних гостей можуть бути певні бонуси, або ж у них за попередні відвідування з’явились вподобання, які теоретично можна віднести до особливих побажань ### Слайд 1

```
requests_logit <- glm(no_of_special_requests ~ required_car_parking_space
                      + log(avg_price_per_room) + no_of_adults + no_of_children
                      + repeated_guest,
                        data = hotel,
                        family = binomial(link = "logit"))

modelsummary(list("basic" = requests_logit),
             gof_omit = "^(?!Num.Obs.|R2 Adj.)",
             stars = TRUE)
```

	basic
(Intercept)	-5.182***
	(0.185)
required_car_parking_space1	1.054***

	(0.071)
log(avg_price_per_room)	0.811***
	(0.041)
no_of_adults	0.641***
	(0.023)
no_of_children	0.310***
	(0.031)
repeated_guest1	0.194*
	(0.079)
Num.Obs.	35674

+ p < 0.1, * p < 0.05, ** p < 0.01, *** p < 0.001

- Як можемо бачити усі коефіцієнти є статистично значущими і жодний не обертається в нуль.З їх знаку можна зробити висновок про додатній статистично значущий зв'язок з бінарною залежною змінною "наявність особливих побажань".
- Додамо до базової моделі контрольні змінні: можна підозрювати, що наявність особливих побажань може бути пов'язана з тривалістю зупинки у готелі. Окрім цього є ще такий фактор як час до прибуття. Оскільки час до прибуття аналогічно до avg_price_per_room варіюється на досить великому проміжку, то було б логічно взяти від lead_price логаритм, аби подивитись на вплив збільшення часу до прибуття на 1%. Також варто врахувати те, що типи кімнат відрізняються між собою як мінімум по кількості дітей, тож у нас є ґрунтовні підстави вважати, що через невідому нам різницю між цими типами кімнат може з'являтися більше чи менше особливих побажань.

```
requests_logit_controls <- glm(no_of_special_requests ~ lead_time + required_car_parking_space
                               + log(avg_price_per_room) + no_of_adults + no_of_children
                               + repeated_guest
                               + no_of_nights
                               + R2 + R3 + R4 + R5 + R6 + R7,
                           data = hotel_room,
                           family = binomial(link = "logit"))

modelsummary(list("basic" = requests_logit, "with_control_vars" = requests_logit_controls),
             gof_omit = "^(?!Num.Obs.)",
             stars = TRUE)
```

	basic	**with_control_vars**
(Intercept)	-5.182***	-5.532***
	(0.185)	(0.208)
required_car_parking_space1	1.054***	1.022***
	(0.071)	(0.071)
log(avg_price_per_room)	0.811***	0.877***
	(0.041)	(0.045)
no_of_adults	0.641***	0.657***
	(0.023)	(0.025)
no_of_children	0.310***	0.545***
	(0.031)	(0.042)
repeated_guest1	0.194*	0.155+
	(0.079)	(0.080)
lead_time		-0.003***
		(0.000)
no_of_nights		0.079***

		(0.006)
R2TRUE		0.475***
		(0.085)
R3TRUE		-1.130
		(1.137)
R4TRUE		0.089**
		(0.033)
R5TRUE		-1.542***
		(0.166)
R6TRUE		-1.082***
		(0.099)
R7TRUE		-0.396+
		(0.221)
Num.Obs.	35674	35674

+ p < 0.1, * p < 0.05, ** p < 0.01, *** p < 0.001

- Додамо фактори взаємодії між змінними. Оскільки в датасеті є записи як з виключно дорослими так і з виключно дітьми, то варто врахувати фактор взаємодії між ними аби включити до розгляду сім'ї з дорослих та дітей. Оскільки як ми побачили раніше коефіцієнт перед required_car_parking_space є статистично значущий, то можна спробувати додати фактор взаємодії між ним та повторним гостем. Окрім того, оскільки більшість записів складають виключно дорослі без дітей, то можна розглянути додатково взаємодію кількості дорослих та повторності гостя.

```
requests_logit_factor <- glm(no_of_special_requests ~ I(log(lead_time+1)) + required_car_parking_space
                                        + log(avg_price_per_room) + no_of_adults + no_of_children
                                        + no_of_adults:no_of_children + no_of_nights + repeated_guest
                                        + repeated_guest:required_car_parking_space
                                        + repeated_guest:no_of_adults
                                        + R2 + R3 + R4 + R5 + R6 + R7,
                      data = hotel_room,
                      family = binomial(link = "logit"))

modelsummary(list("basic" = requests_logit, "with_control_vars" = requests_logit_controls,
                  "all_in_one_with_factors" = requests_logit_factor),
             gof_omit = "^(?!Num.Obs.)",
             stars = TRUE)
```

	basic	with_control_vars	all_in_one_with_factors
(Intercept)	-5.182***	-5.532***	-5.412***
	(0.185)	(0.208)	(0.211)
required_car_parking_space1	1.054***	1.022***	1.081***
	(0.071)	(0.071)	(0.078)
log(avg_price_per_room)	0.811***	0.877***	0.880***
	(0.041)	(0.045)	(0.045)
no_of_adults	0.641***	0.657***	0.702***
	(0.023)	(0.025)	(0.026)
no_of_children	0.310***	0.545***	0.974***
	(0.031)	(0.042)	(0.099)
repeated_guest1	0.194*	0.155+	1.168***
	(0.079)	(0.080)	(0.246)

lead_time		-0.003***	
		(0.000)	
no_of_nights		0.079***	0.081***
		(0.006)	(0.007)
R2TRUE		0.475***	0.329***
		(0.085)	(0.089)
R3TRUE		-1.130	-1.105
		(1.137)	(1.139)
R4TRUE		0.089**	0.122***
		(0.033)	(0.033)
R5TRUE		-1.542***	-1.497***
		(0.166)	(0.166)
R6TRUE		-1.082***	-0.960***
		(0.099)	(0.102)
R7TRUE		-0.396+	-0.278
		(0.221)	(0.221)
l(log(lead_time + 1))			-0.123***
			(0.008)
no_of_adults × no_of_children			-0.248***
			(0.054)
required_car_parking_space1 × repeated_guest1			-0.505*
			(0.210)
no_of_adults × repeated_guest1			-0.777***
			(0.185)
Num.Obs.	35674	35674	35674

+ p < 0.1, * p < 0.05, ** p < 0.01, *** p < 0.001

```
summary(margins(requests_logit_controls))
```

```
##                      factor     AME     SE        z      p   lower   upper
##           avg_price_per_room  0.0021 0.0001  19.9638 0.0000  0.0019  0.0023
##                    lead_time -0.0006 0.0000 -20.2639 0.0000 -0.0007 -0.0006
##                 no_of_adults  0.1495 0.0054  27.6552 0.0000  0.1389  0.1601
##               no_of_children  0.1240 0.0094  13.1984 0.0000  0.1056  0.1424
##                 no_of_nights  0.0179 0.0015  12.2976 0.0000  0.0151  0.0208
##                           R2  0.1087 0.0192   5.6530 0.0000  0.0710  0.1464
##                           R3 -0.2273 0.1857  -1.2235 0.2211 -0.5913  0.1368
##                           R4  0.0203 0.0076   2.6810 0.0073  0.0055  0.0351
##                           R5 -0.2888 0.0219 -13.1955 0.0000 -0.3317 -0.2459
##                           R6 -0.2184 0.0165 -13.2475 0.0000 -0.2507 -0.1861
##                           R7 -0.0877 0.0472  -1.8574 0.0633 -0.1803  0.0048
##              repeated_guest1  0.0354 0.0184   1.9235 0.0544 -0.0007  0.0715
##  required_car_parking_space1  0.2290 0.0146  15.6406 0.0000  0.2003  0.2577
```

Питання №2

2. Чим зумовлено скасування/нескасування резервації?

Можливість з’ясувати чи буде конкретний запис скасований чи нескасований видається досить профітною. Подумки можна швидко прикинути, що можливо повторні гості відмінятимуть записи з меншою ймовірністю. В той же час статус бронювання мав би бути пов’язаний з часом до прибуття (інутїтивно очікується, що чим більше lead_time, то тим менша ймовірність того, що гість все ж прибуде) та, наприклад, кількістю дітей. Хоч записів з дітьми відносно загальної кількості записів не так вже й багато (2559 з 36275), проте захворювання напередодні поїздки або поведінкові проблеми (чи будь-які інші непередбачувані обставини пов’язані з дітьми) мали б зробити свій внесок у ймовірність скасування.

```
nrow(original_hotel %>% filter(no_of_adults != 0 & no_of_children != 0))
```

```
## [1] 2559
```

```
nrow(original_hotel)
```

```
## [1] 36275
```

Побудуємо модель логит, де в ролі залежної бінарної змінної виступатиме атрибут booking_status (‘Canceled’, ‘Not Canceled’)

ДОПИСАТИ

```
denylogit <- glm(booking_status ~ lead_time + no_of_adults + no_of_children
                 + required_car_parking_space + no_of_nights + repeated_guest,
                 family = binomial(link = "logit"),
                 data = hotel)
```

ДОПИСАТИ

```
denylogit_factor <- glm(booking_status ~ lead_time + no_of_adults + no_of_children
                        + required_car_parking_space + no_of_nights + repeated_guest
                        + I(no_of_adults*no_of_children),
                 family = binomial(link = "logit"),
                 data = hotel)
```

ДОПИСАТИ

```
denylogit_with_roomtypes <- glm(booking_status ~ lead_time + no_of_adults
                                + no_of_children + required_car_parking_space
                                + no_of_nights + repeated_guest
                                + R2 + R3 + R4 + R5 + R6 + R7,
                 family = binomial(link = "logit"),
                 data = hotel_room)

coeftest(denylogit, vcov. = vcovHC(denylogit), type = "hc1")
```

```
##
## z test of coefficients:
##
##                               Estimate  Std. Error  z value  Pr(>|z|)
## (Intercept)                 2.24727978  0.05545765  40.5225 < 2.2e-16 ***
## lead_time                  -0.01136218  0.00016911 -67.1899 < 2.2e-16 ***
## no_of_adults               -0.19837831  0.02647863  -7.4920 6.782e-14 ***
## no_of_children             -0.35319585  0.02934264 -12.0369 < 2.2e-16 ***
## required_car_parking_space1  1.30288029  0.08944477  14.5663 < 2.2e-16 ***
## no_of_nights               -0.03988920  0.00778407  -5.1245 2.984e-07 ***
## repeated_guest1             2.25291291  0.22878788   9.8472 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
modelsummary(list("denylogit" = denylogit, "denylogit_factor" = denylogit_factor,
                  "denylogit_with_roomtypes" = denylogit_with_roomtypes),
             stars = TRUE,
             gof_omit = "^(?!Num.Obs.)")
```

	denylogit	denylogit_factor	denylogit_with_roomtypes
(Intercept)	2.247***	2.224***	2.140***
	(0.054)	(0.055)	(0.055)
lead_time	-0.011***	-0.011***	-0.012***
	(0.000)	(0.000)	(0.000)
no_of_adults	-0.198***	-0.185***	-0.109***
	(0.025)	(0.026)	(0.027)
no_of_children	-0.353***	-0.174+	-0.269***
	(0.030)	(0.097)	(0.044)
required_car_parking_space1	1.303***	1.306***	1.318***
	(0.106)	(0.107)	(0.107)
no_of_nights	-0.040***	-0.040***	-0.034***
	(0.007)	(0.007)	(0.007)
repeated_guest1	2.253***	2.260***	2.275***
	(0.265)	(0.265)	(0.266)
I(no_of_adults * no_of_children)		-0.100+	
		(0.051)	
R2TRUE			0.197*
			(0.096)
R3TRUE			-0.262
			(0.955)
R4TRUE			-0.313***
			(0.035)
R5TRUE			-0.127
			(0.155)
R6TRUE			-0.418***
			(0.103)
R7TRUE			-0.109

			(0.217)
Num.Obs.	35674	35674	35674

+ p < 0.1, * p < 0.05, ** p < 0.01, *** p < 0.001

Необхідне для питань 3-4

```
hotel <- hotel %>%
  mutate(log_price = log(avg_price_per_room))


hotel_meal <- hotel %>% mutate(M1 = (type_of_meal_plan == "Meal Plan 1"), M2 = (type_of_meal_plan == "Meal Plan 2
"), M3 = (type_of_meal_plan == "Meal Plan 3"))


hotel_room <- hotel %>% mutate(R2 = (room_type_reserved == "Room_Type 2"), R3 = (room_type_reserved == "Room_Type
3"), R4 = (room_type_reserved == "Room_Type 4"), R5 = (room_type_reserved == "Room_Type 5"), R6 = (room_type_rese
rved == "Room_Type 6"), R7 = (room_type_reserved == "Room_Type 7"))


hotel_market <- hotel %>% mutate(Online = (market_segment_type == "Online"), Corporate = (market_segment_type ==
"Corporate"), Complementary = (market_segment_type == "Complementary"), Aviation = (market_segment_type == "Aviat
ion"))
```

Питання №3

3. Чи є вплив необхідності в паркувальному місці на середню ціну за кімнату?

Під час минулої лабораторної роботи було побудовано довірчі інтервали для середньої ціни для різних категорій людей - тих, кому потрібне, і кому не потрібне паркувальне місце. Було помічено статистично значущу різницю в цінах.

####Тепер, побудуємо базову модель (1) для залежності логарифму ціни за кімнату від необхідності в паркувальному місці. За цією моделлю видно статистично значущий додатний вплив необхідності в паркувальному місці, проте існує багато сумнівів щодо істинності даної моделі. ####Тому, розширимо дану модель додавши декілька логічних контрольних змінних (2). Контрольні змінні - це змінні, які при минулому дослідженні показали зв'язок з необхідністю в паркувальному місці, а саме: кількість людей, наявність спеціальних запитів, кількість ночей і ринковий сегмент. Бачимо, що майже вдвічі знизився коефіцієнт біля основного регресора, тобто дійсно існувала похибка від неврахованих змінних для першої моделі. ####Можемо перевірити модель на стійкість (3)додавши фактор взаємодії паркувального місця і наявністю спец запитів - бачимо, що даний фактор робить коефіцієнт при паркувальному місці незначущим, при цьому для тих, кому необхідне паркувальне місце і при цьому є спеціальні запити коефіцієнт є значущим, і навіть більшим ніж в попередній моделі. ####Крім цього, можемо спробувати зафіксувати ефекти кімнат (4) і часові ефекти на ціну (5). Бачимо в 4 моделі, що знову ключовий коефіцієнт є незначущим, а для 5 моделі він стає значущим, і його інтерпертація наступна: ті, кому необхідне паркувальне місце платять більше в середньому на 4%. Вважаючи, що середня ціна - 100 євро, можемо вважати що це є символічна ціна за аренду паркувального місця.

```
model_car <- feols(log_price ~ required_car_parking_space, data = hotel, vcov = "HC1")
model_car_ext <- feols(log_price ~ required_car_parking_space + no_of_special_requests + no_of_people + no_of_nig
hts + Online + Corporate + Complementary + Aviation, data = hotel_market, vcov = "HC1")
model_car_extvz <- feols(log_price ~ required_car_parking_space*no_of_special_requests + no_of_people + no_of_nig
hts + Online + Corporate + Complementary + Aviation , data = hotel_market, vcov = "HC1")
model_car_room <- feols(log_price ~ required_car_parking_space*no_of_special_requests + no_of_people + no_of_nigh
ts + Online + Corporate + Complementary + Aviation  | room_type_reserved , data = hotel_market)
model_car_room_time <- feols(log_price ~ required_car_parking_space*no_of_special_requests + no_of_people + no_of
_nights + Online + Corporate + Complementary + Aviation  | room_type_reserved + arrival_year_and_month, data = ho
tel_market)


desc_row <- tibble(title = c("Фіксовані ефекти кімнат", "Часові фіксовані ефекти"),
  model_car = c("Hi", "Hi"),
  model_ext = c("Hi", "Hi"),
  model_car_extvz = c("Hi", "Hi"),
  model_car_room = c("Так", "Hi"),
  model_car_room_time = c("Так", "Так"))


modelsummary(list(model_car, model_car_ext, model_car_extvz, model_car_room, model_car_room_time),
             stars = TRUE,
             gof_omit = "^(?!Num.Obs.|R2 Adj.)",
             add_row = desc_row)
```

	(1)	(2)	(3)	(4)	(5)
(Intercept)	4.607***	4.241***	4.242***		
	(0.002)	(0.006)	(0.006)		

required_car_parking_space1	0.112***	0.063***	0.001	0.010	0.027
	(0.011)	(0.008)	(0.011)	(0.006)	(0.014)
no_of_special_requests		0.005	0.002	0.009	0.001
		(0.003)	(0.003)	(0.010)	(0.008)
no_of_people		0.163***	0.163***	0.084***	0.077**
		(0.002)	(0.002)	(0.014)	(0.013)
no_of_nights		-0.018***	-0.018***	-0.020***	-0.018***
		(0.001)	(0.001)	(0.001)	(0.001)
OnlineTRUE		0.159***	0.159***	0.133*	0.137**
		(0.004)	(0.004)	(0.039)	(0.025)
CorporateTRUE		-0.032***	-0.030***	-0.087*	-0.047+
		(0.007)	(0.007)	(0.035)	(0.022)
ComplementaryTRUE		-0.485***	-0.483**	-0.548+	-0.506+
		(0.147)	(0.147)	(0.241)	(0.236)
AviationTRUE		0.271***	0.274***	0.129***	0.077*
		(0.011)	(0.011)	(0.003)	(0.023)
required_car_parking_space1 × no_of_special_requests			0.087***	0.073***	0.060**
			(0.015)	(0.005)	(0.013)
Num.Obs.	35674	35674	35674	35674	35674
R2 Adj.	0.004	0.246	0.247	0.346	0.489
Фіксовані ефекти кімнат	Ні	Ні	Ні	Так	Так
Часові фіксовані ефекти	Ні	Ні	Ні	Ні	Так

+ p < 0.1, * p < 0.05, ** p < 0.01, *** p < 0.001

Питання №4

4. Чи є вплив повторного гостя на середню ціну за кімнату?

При дослідженні, проведеному в минулій роботі було показано, що повторний гість має декілька цікавих особливостей, зокрема, як ми побачили по довірчих інтервалах, він платить менше, причому різниця скаладає приблизно 30% в меншу сторону для повторного гостя. Дуже важко повірити, що різниця настільки велика для повторного гостя. Тому необхідно провести регресійний аналіз.

Спочатку побудуємо базову модель в якій зазначимо лише залежність логарифмованої ціни від повторного гостя, і отримаємо схожий результат. Але у нас є всі сумніви щодо цієї моделі.

Далі побудуємо модель вказавши контрольні змінні, такі як: кількість ночей, кількість людей, ринковий сегмент, необхідність в паркувальному місці - це основні змінні, які можуть корелювати з повторним гостем. Тепер видно вже кращий результат - вплив повторного гостя вже не такий великий, проте значущий. Але і дана модель не викликає повної довіри, бо ми маємо дані панельної природи і, відповідно, потрібно врахувати вплив ефектів кімнат і часу.

По черзі побудуємо відповідні моделі (3 і 4), бачимо, що вплив повторного гостя залишається значущим, але при врахуванні ефектів став дещо менше. Тобто, можемо сказати, що повторний гість дійсно платить менше, але не так багато, як ми вважали спочатку, а всього приблизно на 12%, в це віриться значно більше ніж в попередній результат.

Щоб протестувати модель на стійкість додамо фактор взаємодії між повторним гостем і необхідністю в паркувальному місці, як бачимо коефіцієнт біля повторного гостя несильно змінився, тобто можна сказати, що модель є стійкою і дійсно повторний гость платить дещо менше.

```
model_guest <-  feols(log_price ~ repeated_guest, data = hotel, vcov = "HC1")
model_guest_ext <- feols(log_price ~ repeated_guest + no_of_nights + no_of_people + Online + Corporate + Compleme
ntary + Aviation + required_car_parking_space, data = hotel_market, vcov = "HC1")
model_guest_room <- feols(log_price ~ repeated_guest + no_of_nights + no_of_people + Online + Corporate + Complem
entary + Aviation + required_car_parking_space | room_type_reserved, data = hotel_market)
model_guest_room_time <- feols(log_price ~ repeated_guest + no_of_nights + no_of_people + Online + Corporate + Co
mplementary + Aviation + required_car_parking_space | room_type_reserved + arrival_year_and_month, data = hotel_m
arket)
model_guest_nons <- feols(log_price ~ repeated_guest*required_car_parking_space + no_of_nights + no_of_people + O
nline + Corporate + Complementary + Aviation | room_type_reserved + arrival_year_and_month, data = hotel_market)

desc_row <- tibble(title = c("Фіксовані ефекти кімнат", "Часові фіксовані ефекти"),
  model_guest = c("Ні", "Ні"),
  model_guest_ext = c("Ні", "Ні"),
  model_guest_room = c("Так", "Ні"),
  model_guest_room_time = c("Так", "Так"),
  model_guest_nons = c("Так", "Так"))


modelsummary(list(model_guest, model_guest_ext, model_guest_room, model_guest_room_time, model_guest_nons),
             stars = TRUE,
             gof_omit = "^(?!Num.Obs.|R2 Adj.)",
             add_row = desc_row)
```

	(1)	(2)	(3)	(4)	(5)
(Intercept)	4.617***	4.247***			
	(0.002)	(0.006)			
repeated_guest1	-0.315***	-0.148***	-0.155***	-0.136***	-0.114***
	(0.009)	(0.011)	(0.005)	(0.007)	(0.010)
no_of_nights		-0.018***	-0.021***	-0.018***	-0.018***
		(0.001)	(0.001)	(0.001)	(0.001)
no_of_people		0.162***	0.083***	0.075**	0.075**
		(0.002)	(0.014)	(0.013)	(0.013)
OnlineTRUE		0.160***	0.136**	0.137***	0.136***
		(0.003)	(0.035)	(0.022)	(0.022)
CorporateTRUE		0.009	-0.046	-0.011	-0.010
		(0.008)	(0.031)	(0.019)	(0.019)
ComplementaryTRUE		-0.471**	-0.533+	-0.494+	-0.481+
		(0.147)	(0.246)	(0.241)	(0.243)
AviationTRUE		0.287***	0.142***	0.089***	0.088**
		(0.012)	(0.012)	(0.014)	(0.017)
required_car_parking_space1		0.074***	0.075***	0.080***	0.097***
		(0.007)	(0.003)	(0.004)	(0.005)
repeated_guest1 × required_car_parking_space1					-0.153***
					(0.024)
Num.Obs.	35674	35674	35674	35674	35674
R2 Adj.	0.024	0.250	0.350	0.493	0.493
Фіксовані ефекти кімнат	Ні	Ні	Так	Так	Так
Часові фіксовані ефекти	Ні	Ні	Ні	Так	Так

+ p < 0.1, * p < 0.05, ** p < 0.01, *** p < 0.001

Питання №5

##(Залишається тільки в звіті) #### 5. Чи є вплив кількості людей на середню ціну за кімнату?

Ми вже неодноразово бачили що збільшення кількості дорослих і дітей має зв'язок зі збільшенням ціни за кімнату, але чи можна вважати, що він причинно-наслідковий?

Спочатку побудуємо примітивну модель (1) і на ній бачимо, що збільшення дорослого на 1 призведе до підвищення ціни на 16%, а збільшення дитини на 1 аж на 25%. У нас немає приводу довіряти цій моделі.

Розширимо дану модель додавши контрольні змінні (2), які можуть корелювати з кількістю людей - це кількість ночей і сегмент ринку, також додамо фактор взаємодії між дітьми і дорослими, бо логічно припустити, що існують зміни для кожного з цих регресорів в залежності від іншого. Бачимо, що значення коефіцієнтів зменшилися, особливо для кількості дітей. Але знову є всі здогадки, що модель не є ідеальною.

Моделі 3 і 4 відповідно додають фіксацію щодо впливу типу кімнати і часового проміжку. Бачимо, що коефіцієнти знову змінилися - дещо зросли для дітей і впали для дорослих, тут ми вже врахували всі доступні нам ефекти і можемо вважати, що збільшення кількості дітей і дорослих на одиницю має зв'язок зі збільшенняи ціни за кімнату на 6 і 8 відсотків відповідно.

Протестуємо дану модель на стійкість (5, 6): спочатку (5) перетворимо кількість дітей на бінарну змінну там де їх багато(більше ніж 1) і мало, бачимо, що коефіцієнт при кількості дорослих не сильно змінився, тобто модель є стійкою відносно кількості дорослих. Далі (6) зробимо схоже перетворення кількості дорослих на бінарну змінну - там де їх багато(більше ніж 2) і мало, коефіцієнт біля кількості дітей значно змінився, тобто модель не стійка щодо цієї змінної і могли б існувати інші змінні для зменшення OVB.

```
model_people <- feols(log_price ~ no_of_adults + no_of_children, data = hotel, vcov = "HC1")
model_people_ext <- feols(log_price ~ no_of_adults*no_of_children + no_of_nights + Online + Corporate + Complemen
tary + Aviation  , data = hotel_market, vcov = "HC1")
model_people_room <- feols(log_price ~ no_of_adults*no_of_children + no_of_nights + Online + Corporate + Compleme
ntary + Aviation |room_type_reserved   , data = hotel_market)
model_people_time <- feols(log_price ~ no_of_adults*no_of_children + no_of_nights + Online + Corporate + Compleme
ntary + Aviation |room_type_reserved + arrival_year_and_month , data = hotel_market)
#стійкість
model_people_st <- feols(log_price ~ no_of_adults*no_of_children + no_of_nights + Online + Corporate + Complement
ary + Aviation |room_type_reserved + arrival_year_and_month , data = hotel_market %>% mutate (no_of_children = if
else(no_of_children > 1, 1, 0)))
model_people_st_2 <- feols(log_price ~ no_of_adults*no_of_children + no_of_nights + Online + Corporate + Compleme
ntary + Aviation |room_type_reserved + arrival_year_and_month , data = hotel_market %>% mutate (no_of_adults = if
else(no_of_adults > 2, 1, 0)))

modelsummary(list(model_people, model_people_ext, model_people_room, model_people_time, model_people_st, model_pe
ople_st_2),
             stars = TRUE,
             gof_omit = "^(?!Num.Obs.|R2 Adj.)"
             )
```

	(1)	(2)	(3)	(4)	(5)	(6)
(Intercept)	4.281***	4.316***				
	(0.006)	(0.006)				
no_of_adults	0.164***	0.119***	0.072+	0.061+	0.062+	0.215***
	(0.003)	(0.003)	(0.030)	(0.029)	(0.026)	(0.012)
no_of_children	0.246***	0.051***	0.078*	0.048	0.102+	0.121*
	(0.004)	(0.014)	(0.030)	(0.027)	(0.043)	(0.037)
no_of_nights		-0.017***	-0.021***	-0.018***	-0.018***	-0.017***
		(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
OnlineTRUE		0.163***	0.137**	0.139**	0.144***	0.143***
		(0.003)	(0.037)	(0.024)	(0.021)	(0.023)
CorporateTRUE		-0.049***	-0.090+	-0.050	-0.052	-0.078**
		(0.007)	(0.042)	(0.030)	(0.028)	(0.015)
ComplementaryTRUE		-0.510***	-0.558+	-0.518+	-0.503+	-0.526+
		(0.151)	(0.234)	(0.232)	(0.224)	(0.223)

AviationTRUE		0.241***	0.116***	0.063***	0.064**	0.036
		(0.011)	(0.016)	(0.009)	(0.011)	(0.027)
no_of_adults × no_of_children		0.092***	0.026	0.040	0.031	-0.061
		(0.007)	(0.034)	(0.031)	(0.032)	(0.058)
Num.Obs.	35674	35674	35674	35674	35674	35674
R2 Adj.	0.179	0.257	0.346	0.490	0.481	0.504

+ p < 0.1, * p < 0.05, ** p < 0.01, *** p < 0.001

```
linearHypothesis(model_people_st, c("no_of_children = 0", "no_of_adults:no_of_children = 0"))
```

```
## Linear hypothesis test
##
## Hypothesis:
## no_of_children = 0
## no_of_adults:no_of_children = 0
##
## Model 1: restricted model
## Model 2: log_price ~ no_of_adults * no_of_children + no_of_nights + Online +
##     Corporate + Complementary + Aviation | room_type_reserved +
##     arrival_year_and_month
##
##   Res.Df Df  Chisq Pr(>Chisq)
## 1  35644
## 2  35642  2 19.339  6.319e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
linearHypothesis(model_people_st_2, c("no_of_adults = 0", "no_of_adults:no_of_children = 0"))
```

```
## Linear hypothesis test
##
## Hypothesis:
## no_of_adults = 0
## no_of_adults:no_of_children = 0
##
## Model 1: restricted model
## Model 2: log_price ~ no_of_adults * no_of_children + no_of_nights + Online +
##     Corporate + Complementary + Aviation | room_type_reserved +
##     arrival_year_and_month
##
##   Res.Df Df  Chisq Pr(>Chisq)
## 1  35644
## 2  35642  2 431.46  < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

МУСОРКА

дослідимо наскільки більша ймовірність відмови бронювання для повторних гостей порівняно з неповторними

```
numeric_means <- hotel %>%
  dplyr::select(lead_time, avg_price_per_room, no_of_adults, no_of_children, no_of_nights) %>%
  summarize(across(everything(), ~ mean(.)))

new_data <- cbind(numeric_means,
                  data.frame(factorno_of_special_requests = c(0, 1),
                             required_car_parking_space = factor(0, levels = levels(hotel$required_car_parking_sp
ace)),
                             repeated_guest = factor(0, levels = levels(hotel$repeated_guest))))

new_data <- rbind(new_data, new_data)
new_data[2, "factorno_of_special_requests"] <- 1

prediction_logit <- predict(requests_logit, new_data, type = "response")

difference <- prediction_logit[2] - prediction_logit[1]
difference
```

```
## 2
## 0
```

кореляції (пріколи)