


Груповий проект з аналізу даних  
(Лаб **2**. Статистичне виведення)

# Бронювання в готелі

## Склад команди:

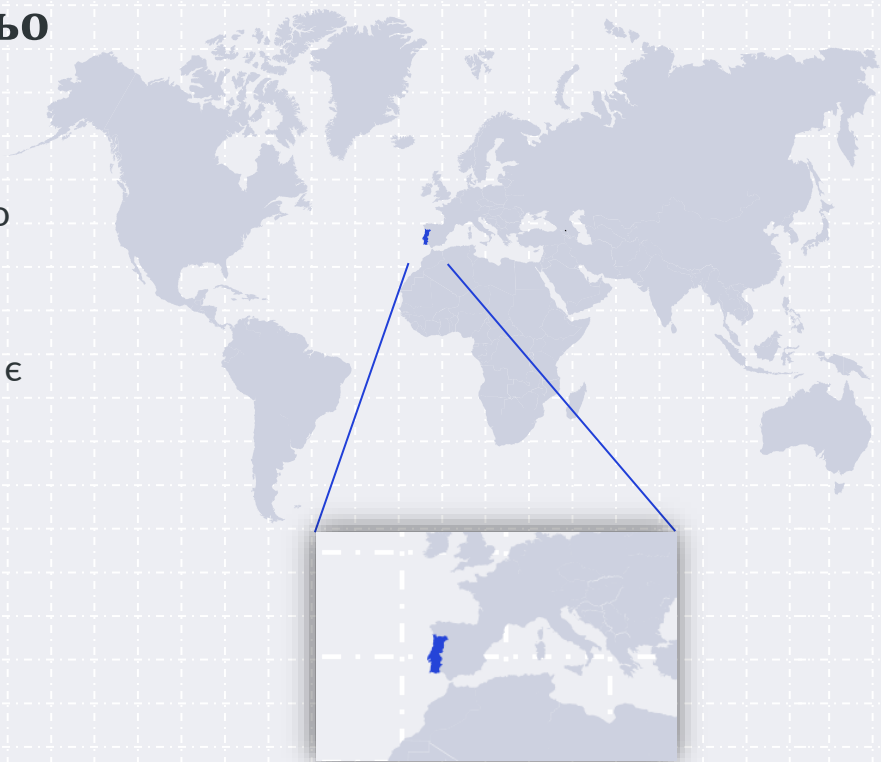
- Пономаренко Олександр (КМ-12)
  - Борисенко Данило (КМ-11)
  - Земляний Даниїл (КМ-12)
  - Лук'яненко Василь (КМ-13)
  - Заїченко Дамир (КМ-13)
- 

## Опис даних та як їх було попередньо опрацьовано:

У цього датасету є більший "попередник", в описі якого вказано, що автор датасету брав натхнення з дослідження, яке використовувало дані, зібрані у двох дійсно існуючих готелях Португалії. Поточний датасет є очищеною версією, що не містить майже незадіяних змінних та купи пропущених даних.

Опрацювання складалося з наступних пунктів:

- Перевірка пропущених значень
- Переведення відповідних змінних у фактор
- Пошук викидів



# Змінні та їх опис

**id:** унікальний ідентифікатор кожного бронювання

**no\_of\_adults:** Кількість дорослих

**no\_of\_children:** Кількість дітей

**no\_of\_weekend\_nights:** Кількість ночей у вихідні (субота або неділя), які гості провели або забронювали для перебування в готелі

**no\_of\_week\_nights:** Кількість ночей у будні (з понеділка по п'ятницю), які гості провели або забронювали для перебування в готелі

**type\_of\_meal\_plan:** Тип харчування, заброньований клієнтом

**required\_car\_parking\_space:** Чи потрібне клієнту місце для паркування автомобіля? (0 - Ні, 1 - Так)

**room\_type\_reserved:** Тип заброньованого номера. Значення закодовані INN готелями.

**lead\_time:** Кількість днів між датою бронювання та датою прибуття

**arrival\_year:** Рік прибуття

**arrival\_month:** Місяць прибуття

**arrival\_date:** День місяця прибуття

**market\_segment\_type:** Призначення ринкового сегменту

**repeated\_guest:** Чи є клієнт постійним гостем? (0 - Ні, 1 - Так)

**no\_of\_previous\_cancellations:** Кількість попередніх бронювань, скасованих клієнтом перед поточним бронюванням

**no\_of\_previous\_bookings\_not\_canceled:** Кількість попередніх бронювань, не скасованих клієнтом перед поточним бронюванням

**avg\_price\_per\_room:** Середня ціна за номер на день; ціни на номери динамічні. (в євро)

**no\_of\_special\_requests:** Загальна кількість особливих бажань, зроблених клієнтом (наприклад, високий поверх, вид з номера тощо)

# Пригадаємо структуру датасету

```
'data.frame':  35674 obs. of  22 variables:
 $ Booking_ID           : chr  "INN00001" "INN00002" "INN00003" "INN00004" ...
 $ no_of_adults         : int   2 2 1 2 2 2 2 2 3 2 ...
 $ no_of_children       : int   0 0 0 0 0 0 0 0 0 0 ...
 $ no_of_weekend_nights : int   1 2 2 0 1 0 1 1 0 0 ...
 $ no_of_week_nights    : int   2 3 1 2 1 2 3 3 4 5 ...
 $ type_of_meal_plan     : chr  "Meal Plan 1" "Not Selected" "Meal Plan 1" "Meal Plan 1" ...
 $ required_car_parking_space : Factor w/ 2 levels "0","1": 1 1 1 1 1 1 1 1 1 1 ...
 $ room_type_reserved   : Factor w/ 7 levels "Room_Type 1",...: 1 1 1 1 1 1 1 4 1 4 ...
 $ lead_time            : int  224 5 1 211 48 346 34 83 121 44 ...
 $ arrival_year         : int  2017 2018 2018 2018 2018 2018 2018 2017 2018 2018 ...
 $ arrival_month        : int  10 11 2 5 4 9 10 12 7 10 ...
 $ arrival_date         : int   2 6 28 20 11 13 15 26 6 18 ...
 $ market_segment_type  : chr  "offline" "Online" "Online" "Online" ...
 $ repeated_guest       : Factor w/ 2 levels "0","1": 1 1 1 1 1 1 1 1 1 1 ...
 $ no_of_previous_cancellations : int   0 0 0 0 0 0 0 0 0 0 ...
 $ no_of_previous_bookings_not_canceled: int   0 0 0 0 0 0 0 0 0 0 ...
 $ avg_price_per_room   : num   65 106.7 60 100 94.5 ...
 $ no_of_special_requests : num   0 1 0 0 0 1 1 1 1 1 ...
 $ booking_status       : Factor w/ 2 levels "Canceled","Not_Canceled": 2 2 1 1 1 1 2 2 2 2 ...
 $ arrival_year_and_month : chr  "2017-10-2" "2018-11-6" "2018-2-28" "2018-5-20" ...
 $ no_of_people         : int   2 2 1 2 2 2 2 2 3 2 ...
 $ no_of_nights         : int   3 5 3 2 2 2 4 4 4 5 ...
```

# Дослідницькі питання



01

Як змінюється кількість дорослих та дітей в залежності від типу номеру та плану харчування

02

Яким чином розподіляється кількість особливих побажань в залежності від типу номеру?

03

Як впливає кількість дорослих і дітей на скасування бронювання?

04

Що впливає на кількість проведених вихідних та робочих ночей у готелі?

05

Як різні типи кімнат впливають на кількість попередніх скасувань/бронювань?

06

Чи є різниця в кількості попередніх скасувань для клієнтів, які вимагають паркувальне місце і тих, хто його не потребує?



# Оновлені дослідницькі питання



01

Що впливає на вибір типу номеру чи плану харчування?

02

Чи є істотною наявність особливих побажань?

03

Які характерні риси скасованих записів?

04

Що впливає на кількість проведених ночей у готелі?

05

Що впливає на кількість попередніх скасувань/ не скасувань?

06

Які характерні риси бронювань з потребою у паркувальному місці?

07

Які нові цікаві відомості про повторних гостей?



# Гіпотези



01

- 1) Середня кількість людей в різних кімнатах однакова
- 2) Середня ціна різних кімнат однакова

03

- 6) Не відмінені бронювання в середньому мають більше людей, ніж відмінені
- 7) В середньому на нескасованих бронюваннях час до прибуття є більшим, ніж у скасованих

06

- 10) Немає різниці в середніх кількостях людей для записів, що потребують і не потребують паркувальне місце
- 11) Люди, яким не потрібно паркувальне місце в середньому платять більше за тих, кому паркувальне місце потрібне

02

- 3) Середня кількість людей не залежить від особливих бажань
- 4) Середня кількість проведених ночей не залежить від особливих бажань
- 5) Ті хто не має особливих бажань платять більше

05

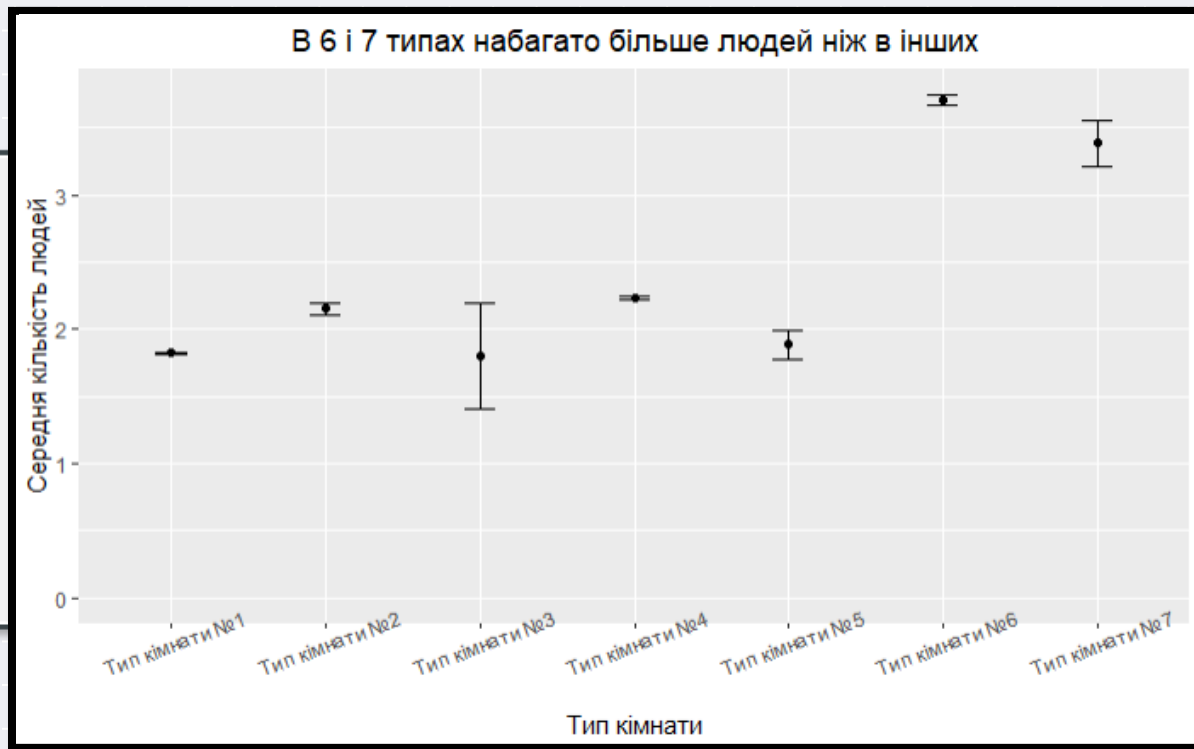
- 8) Відмінені і невідмінені бронювання мають однакову кількість попередніх скасування бронювань
- 9) Відмінені і невідмінені бронювання мають однакову кількість попередніх НЕ скасування бронювань

07

- 12) Тих гостей, які приїжджають вперше і не вперше однакова кількість
- 13) Рівність (в середньому) оплати для гостей, що прибули вперше і не вперше
- 14) незалежність записів з повторним гостем від інших величин, таких як статус бронювання і необхідність в паркувальному місці.



## Що впливає на вибір типу номеру чи плану харчування?

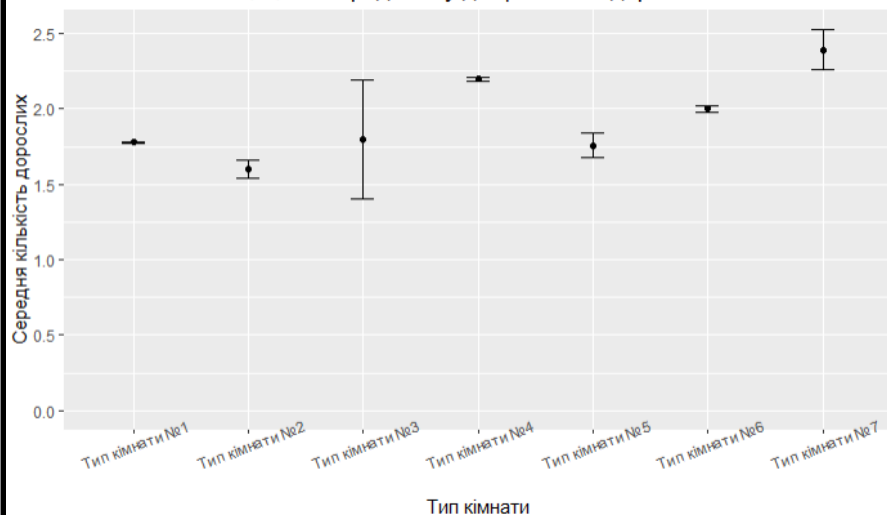




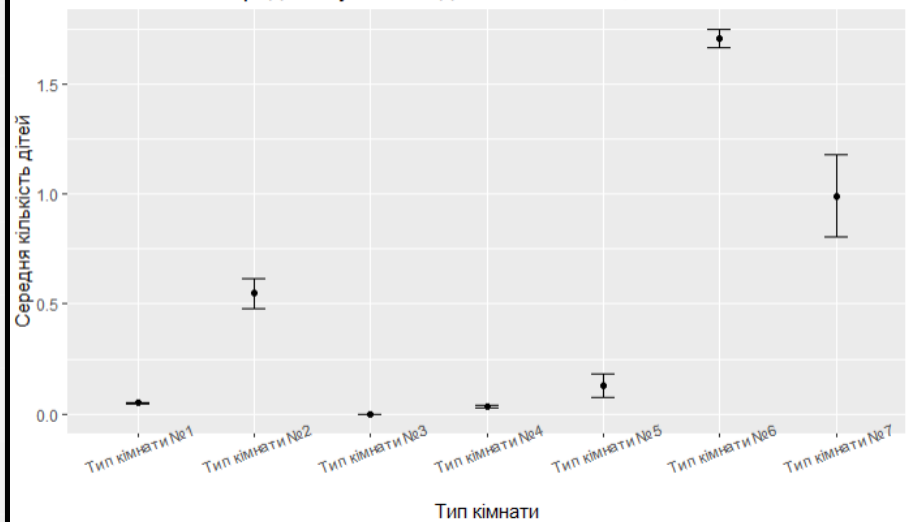
# Що впливає на вибір типу номеру чи плану харчування?

В 6 і 7 типах набагато більше людей ніж в інших

В кімнатах 4, 6, 7 в середньому дещо більше дорослих ніж в інших

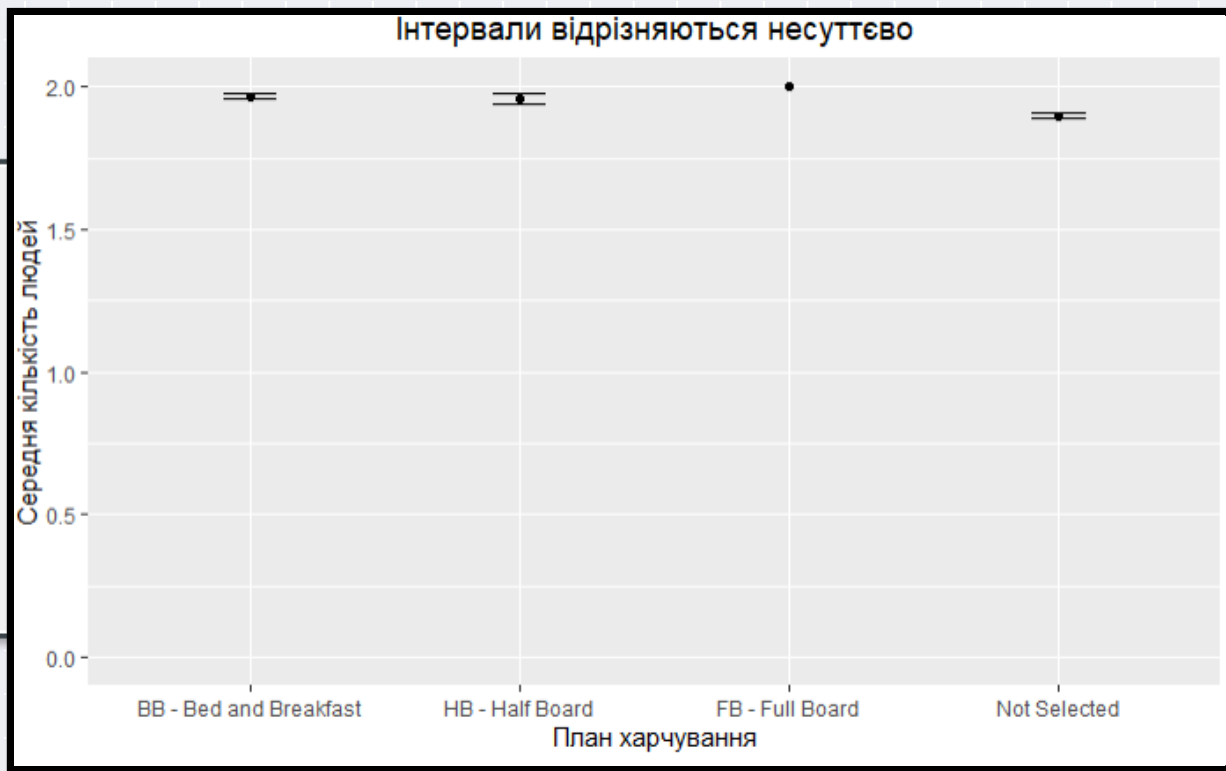


В 2 і 7 в середньому більше дітей ніж в інших, в 6 - значно більше



Тип кімнати №1 Тип кімнати №2 Тип кімнати №3 Тип кімнати №4 Тип кімнати №5 Тип кімнати №6 Тип кімнати №7  
Тип кімнати

## Що впливає на вибір типу номеру чи плану харчування?

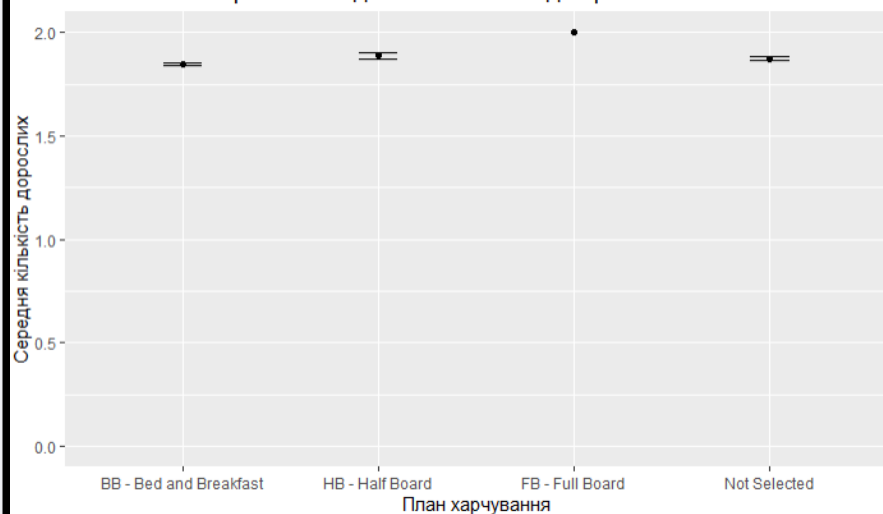


01

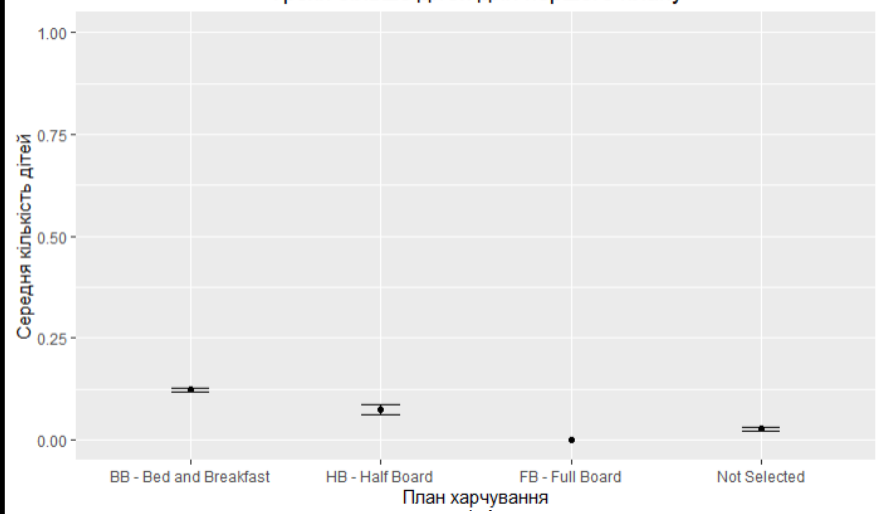
# Що впливає на вибір типу номеру чи плану харчування?

Інтервали відрізняються не суттєво

Приблизно однакова кількість для різних планів



Трохи більше дітей для першого плану



BB - Bed and Breakfast

HB - Half Board

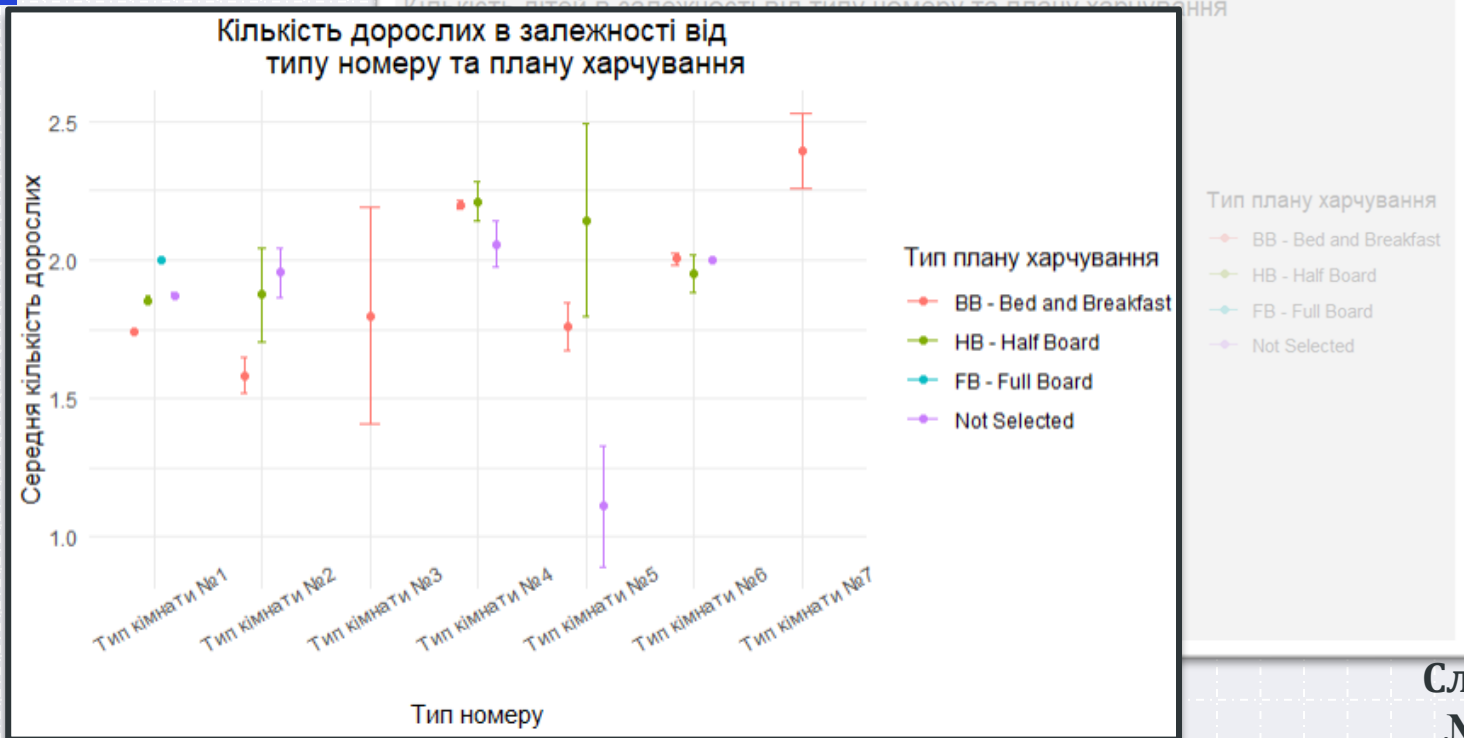
FB - Full Board

Not Selected

План харчування

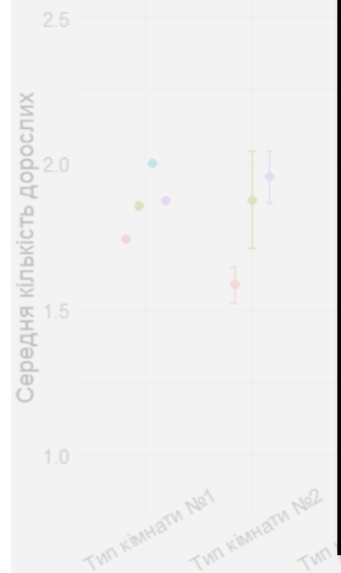
Слайд  
№11

## Що впливає на вибір типу номеру чи плану харчування?

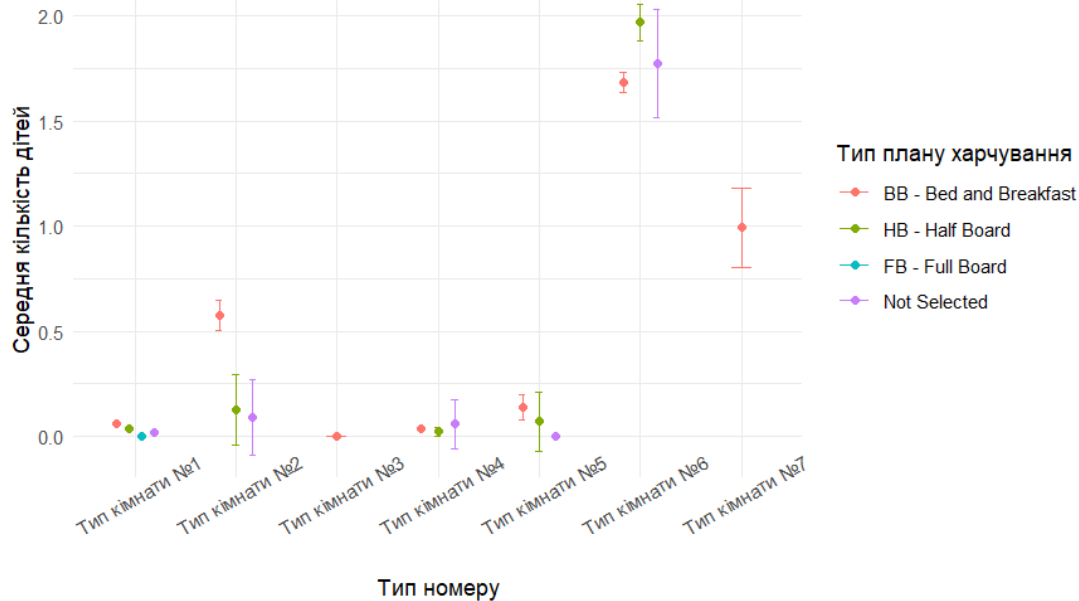


## Що впливає на вибір типу номеру чи плану харчування?

Кількість дорослих в зале



Кількість дітей в залежності від типу номеру та плану харчування



## Множинне тестування

1) Протестуємо гіпотезу  $H_0$ : середня кількість людей в різних кімнатах однакова

```
data: hotel_no_people$no_of_people and factor(hotel$room_type_reserved)
```

|             | Room_Type 1 | Room_Type 2 | Room_Type 3 | Room_Type 4 | Room_Type 5 | Room_Type 6 |
|-------------|-------------|-------------|-------------|-------------|-------------|-------------|
| Room_Type 2 | < 2e-16     | -           | -           | -           | -           | -           |
| Room_Type 3 | 0.90231     | 0.17804     | -           | -           | -           | -           |
| Room_Type 4 | < 2e-16     | 0.00087     | 0.11797     | -           | -           | -           |
| Room_Type 5 | 0.29180     | 1.3e-05     | 0.72928     | 1.5e-09     | -           | -           |
| Room_Type 6 | < 2e-16     | < 2e-16     | 0.00087     | < 2e-16     | < 2e-16     | -           |
| Room_Type 7 | < 2e-16     | < 2e-16     | 0.00069     | < 2e-16     | < 2e-16     | 0.00091     |

P value adjustment method: BH

# 01

## Множинне тестування

### 2) Протестуємо гіпотезу $H_0$ : середня ціна в різних кімнатах однакова

data: hotel\$avg\_price\_per\_room and factor(hotel\$room\_type\_reserved)

|             | Room_Type 1 | Room_Type 2 | Room_Type 3 | Room_Type 4 | Room_Type 5 | Room_Type 6 |
|-------------|-------------|-------------|-------------|-------------|-------------|-------------|
| Room_Type 2 | 4.1e-05     | -           | -           | -           | -           | -           |
| Room_Type 3 | 0.68270     | 0.49608     | -           | -           | -           | -           |
| Room_Type 4 | < 2e-16     | < 2e-16     | 0.16471     | -           | -           | -           |
| Room_Type 5 | < 2e-16     | < 2e-16     | 0.10110     | 0.01586     | -           | -           |
| Room_Type 6 | < 2e-16     | < 2e-16     | 0.00419     | < 2e-16     | < 2e-16     | -           |
| Room_Type 7 | < 2e-16     | < 2e-16     | 0.00092     | < 2e-16     | < 2e-16     | 2.6e-07     |

P value adjustment method: BH

data: hotel\_no\_people\$no\_of\_people and factor(hotel\$room\_type\_reserved)

|             | Room_Type 1 | Room_Type 2 | Room_Type 3 | Room_Type 4 | Room_Type 5 | Room_Type 6 |
|-------------|-------------|-------------|-------------|-------------|-------------|-------------|
| Room_Type 2 | < 2e-16     | -           | -           | -           | -           | -           |
| Room_Type 3 | 0.90231     | 0.17804     | -           | -           | -           | -           |
| Room_Type 4 | < 2e-16     | 0.00087     | 0.11797     | -           | -           | -           |
| Room_Type 5 | 0.29180     | 1.3e-05     | 0.72928     | 1.5e-09     | -           | -           |
| Room_Type 6 | < 2e-16     | < 2e-16     | 0.00087     | < 2e-16     | < 2e-16     | -           |
| Room_Type 7 | < 2e-16     | < 2e-16     | 0.00069     | < 2e-16     | < 2e-16     | 0.00091     |

P value adjustment method: BH

### 1) Протестуємо гіпотезу $H_0$ : середня кількість людей в різних кімнатах однакова

# 01

## Висновки

- **6 та 7** типи кімнат найчастіше бронюють сім'ї з **1-3** дітьми
- загалом ніякому конкретному плану харчування перевагу не надають
- **Full-Board** з певних причин умисно уникають. На **36** тисяч записів існує тільки одне бронювання, де було обрано дану опцію
- в **5-ому** типі кімнат значно менша кількість дорослих вирішує не обирати план харчування. При цьому серед всіх резервацій у **5-ий** тип кімнат в яких є діти жодного разу не було обрано опцію **Not Selected**
- з відносно високою ймовірністю можна припускати, що у **1 та 3** типи кімнат в середньому селиться приблизно однакова кількість осіб



# 02

## Чи є істотною наявністю особливих побажань?

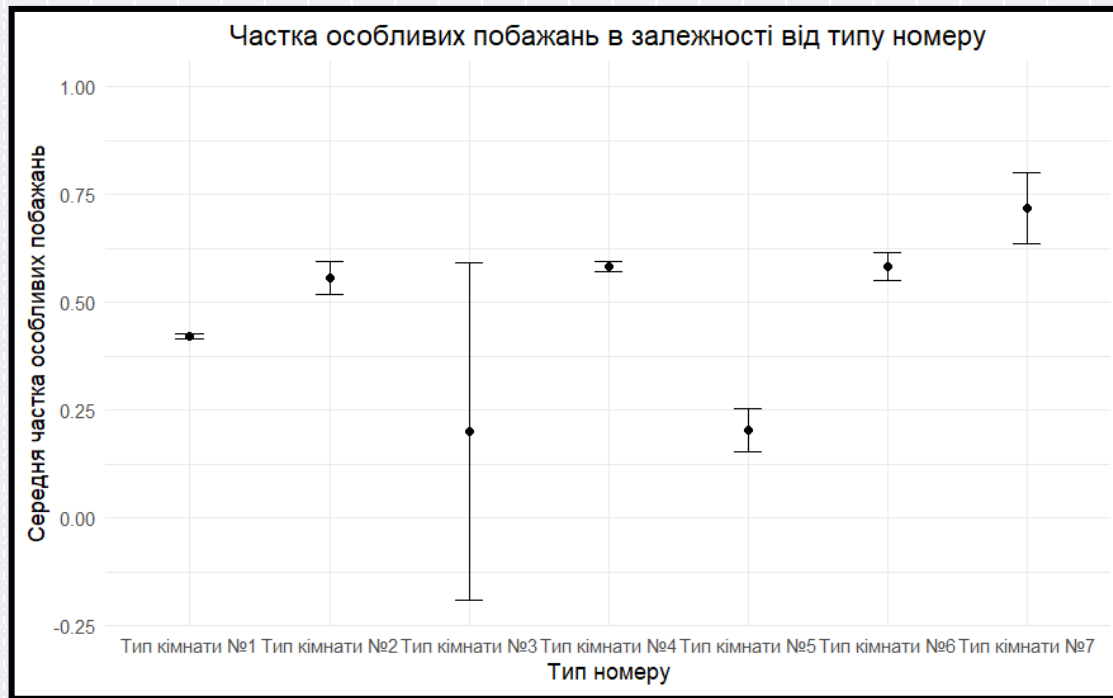
Побудуємо довірчі інтервали для середнього числа особливих побажань для кожного типу номеру

| room_type_reserved<br><fctr> | mean_requests<br><dbl> | sd_requests<br><dbl> | n_requests<br><int> | ci_low_requests<br><dbl> | ci_high_requests<br><dbl> |
|------------------------------|------------------------|----------------------|---------------------|--------------------------|---------------------------|
| Room_Type 1                  | 0.4210127              | 0.4937305            | 27707               | 0.4151992                | 0.4268263                 |
| Room_Type 2                  | 0.5565749              | 0.4971692            | 654                 | 0.5184715                | 0.5946783                 |
| Room_Type 3                  | 0.2000000              | 0.4472136            | 5                   | -0.1919928               | 0.5919928                 |
| Room_Type 4                  | 0.5818182              | 0.4933015            | 5995                | 0.5693310                | 0.5943054                 |
| Room_Type 5                  | 0.2024291              | 0.4026262            | 247                 | 0.1522178                | 0.2526405                 |
| Room_Type 6                  | 0.5827187              | 0.4933702            | 949                 | 0.5513289                | 0.6141084                 |
| Room_Type 7                  | 0.7179487              | 0.4519337            | 117                 | 0.6360589                | 0.7998385                 |

# 02

Чи є істотною наявністю особливих побажань?

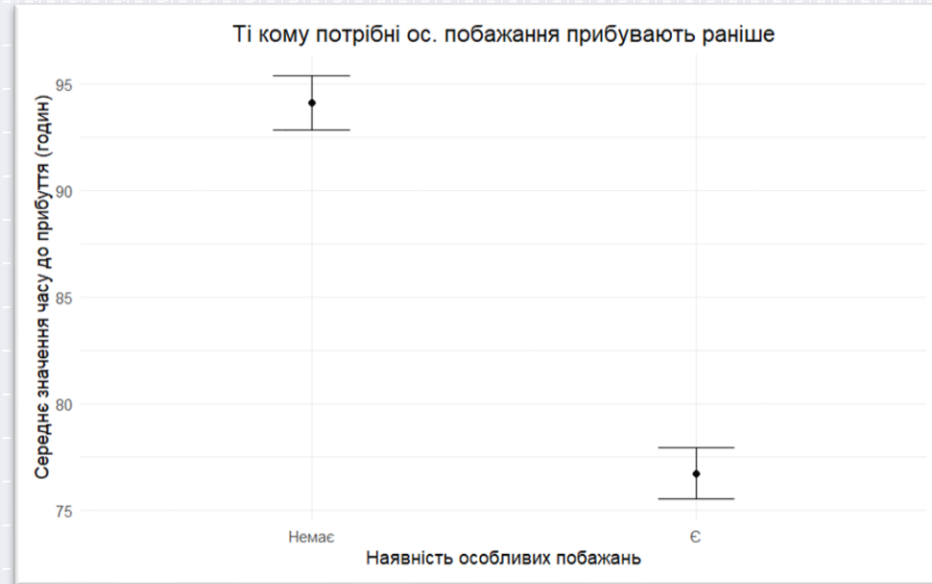
Побудуємо довірчі інтервали для середнього числа особливих побажань для кожного типу номеру



# 02

Чи є істотною наявністю особливих побажань?

Побудуємо довірчі інтервали для середнього значення часу до прибуття, в залежності від наявності особливих бажань

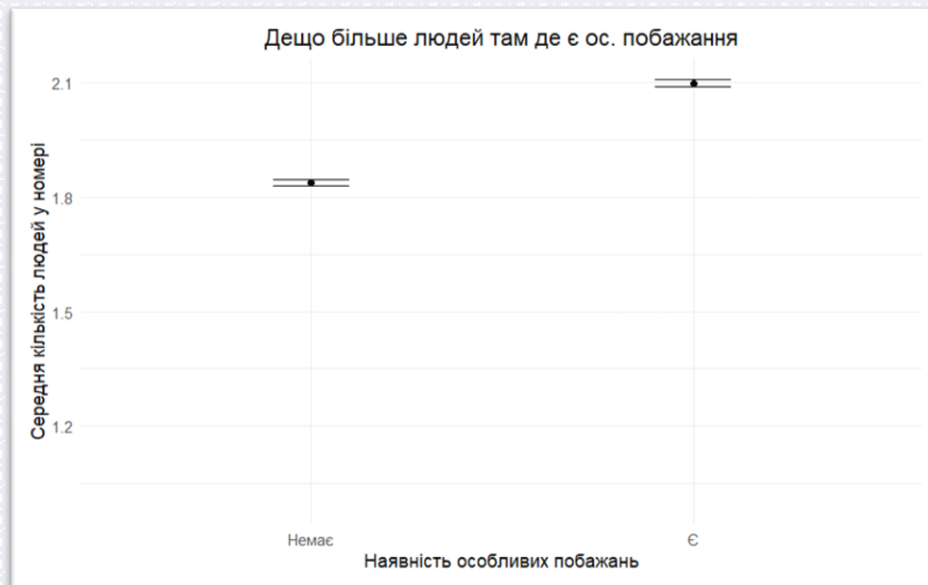


| no_of_special_requests<br><dbl> | mean_lead_time<br><dbl> | sd_lead_time<br><dbl> | n_requests<br><int> | ci_low_requests<br><dbl> | ci_high_requests<br><dbl> |
|---------------------------------|-------------------------|-----------------------|---------------------|--------------------------|---------------------------|
| 0                               | 94.13488                | 91.04175              | 19469               | 92.85604                 | 95.41372                  |
| 1                               | 76.72422                | 78.36984              | 16205               | 75.51760                 | 77.93085                  |

# 02

Чи є істотною наявністю особливих побажань?

Побудуємо тепер довірчі інтервали для середньої кількості людей у кімнаті, в залежності від наявності особливих побажань



| no_of_special_requests<br><dbl> | mean_no_of_people<br><dbl> | sd_no_of_people<br><dbl> | n_requests<br><int> | ci_low_requests<br><dbl> | ci_high_requests<br><dbl> |
|---------------------------------|----------------------------|--------------------------|---------------------|--------------------------|---------------------------|
| 0                               | 1.837639                   | 0.6272597                | 19469               | 1.828828                 | 1.846450                  |
| 1                               | 2.098858                   | 0.6313827                | 16205               | 2.089137                 | 2.108579                  |

Слайд  
№20

## 02

### Формулювання гіпотез: Проведемо тест Волда та **t-test** Велча на рівні значущості **0.05**

**Нульова гіпотеза: "У записах в яких не потрібні і потрібні особливі  
бажання, в середньому заселялась однакова кількість людей"**

```
Mean (Requests): 2.098858  
Mean (No Requests): 1.837639  
T-statistic: -39.02298  
P-value: 0  
95% Confidence Interval: -0.274339 -0.2480991
```

Welch Two Sample t-test

```
data: no_of_people by no_of_special_requests  
t = -39.023, df = 34425, p-value < 2.2e-16  
alternative hypothesis: true difference in means between group 0 and group 1 is not equal to 0  
95 percent confidence interval:  
-0.2743395 -0.2480986  
sample estimates:  
mean in group 0 mean in group 1  
1.837639 2.098858
```

# 02

Чи є істотною наявність особливих побажань?

Протестуємо гіпотезу про те, що люди, яким не потрібні і потрібні спец запити ночують в середньому однакову кількість ночей

```
Mean (Requests): 3.166739
Mean (No Requests): 2.934203
T-statistic: -12.25289
P-value: 1.621207e-34
95% Confidence Interval: -0.2697318 -0.1953393
```

Welch Two Sample t-test

```
data: no_of_nights by no_of_special_requests
t = -12.253, df = 34006, p-value < 2.2e-16
alternative hypothesis: true difference in means between group 0 and group 1 is not
equal to 0
95 percent confidence interval:
-0.2697331 -0.1953380
sample estimates:
mean in group 0 mean in group 1
2.934203          3.166739
```

| mean<br><dbl> | sd<br><dbl> | n<br><int> | a<br><dbl> | b<br><dbl> |
|---------------|-------------|------------|------------|------------|
| 2.934203      | 1.74762     | 19469      | 2.909655   | 2.958752   |
| mean<br><dbl> | sd<br><dbl> | n<br><int> | a<br><dbl> | b<br><dbl> |
| 3.166739      | 1.815031    | 16205      | 3.138793   | 3.194684   |

# 02

Чи є істотною наявність особливих побажань?

Тепер протестуємо гіпотезу про те, що люди, яким НЕ потрібні спеціальні запити, в середньому платять більше

```
Mean (Requests): 111.3596
Mean (No Requests): 99.98524
T-statistic: -32.9086
P-value: 8.278696e-238
95% Confidence Interval: -12.05183 -10.69696
```

Welch Two Sample t-test

```
data: avg_price_per_room by no_of_special_requests
t = -32.909, df = 33178, p-value < 2.2e-16
alternative hypothesis: true difference in means between group 0 and group 1 is less
than 0
95 percent confidence interval:
      -Inf -10.80586
sample estimates:
mean in group 0 mean in group 1
    99.98524      111.35964
```

| mean<br><dbl> | sd<br><dbl> | n<br><int> | a<br><dbl> | b<br><dbl> |
|---------------|-------------|------------|------------|------------|
| 99.98524      | 30.85209    | 19469      | 99.55187   | 100.4186   |
| mean<br><dbl> | sd<br><dbl> | n<br><int> | a<br><dbl> | b<br><dbl> |
| 111.3596      | 33.81783    | 16205      | 110.839    | 111.8803   |

## 02

# Чи є істотною наявність особливих побажань?

**Таким чином отримуємо проміжні висновки:**

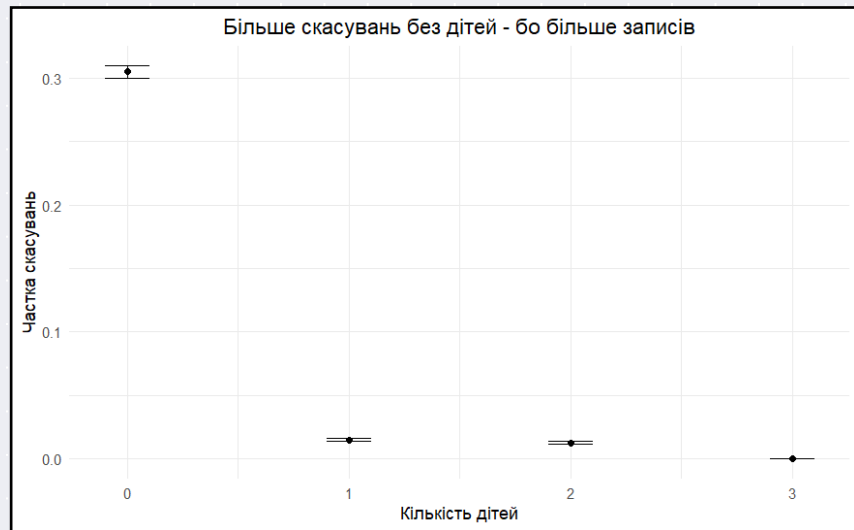
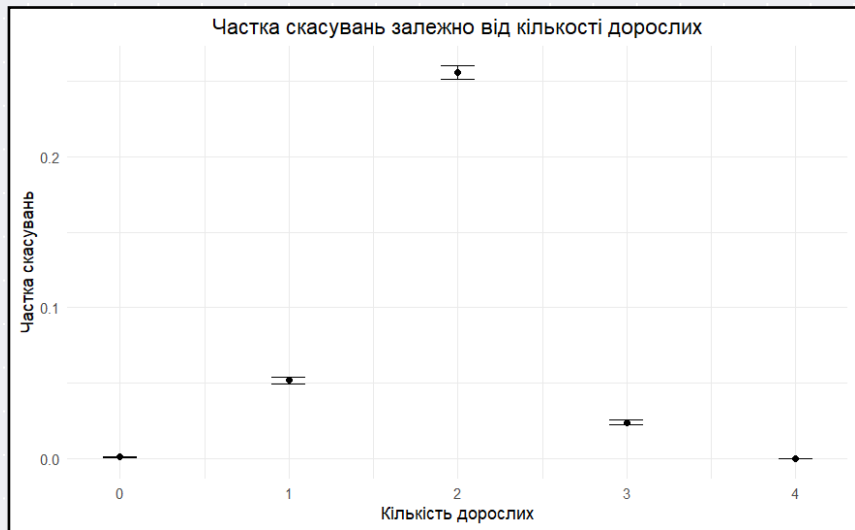
**За перевіреними гіпотезами можемо спостерігати, що наявність особливих бажань виявилась дійсно істотною. У записах, де вимагались особливі бажання, було в середньому більше людей, які в середньому заселялись раніше, жили трохи довше, і платили трохи більше**



# 03

## Які характерні риси скасованих записів?

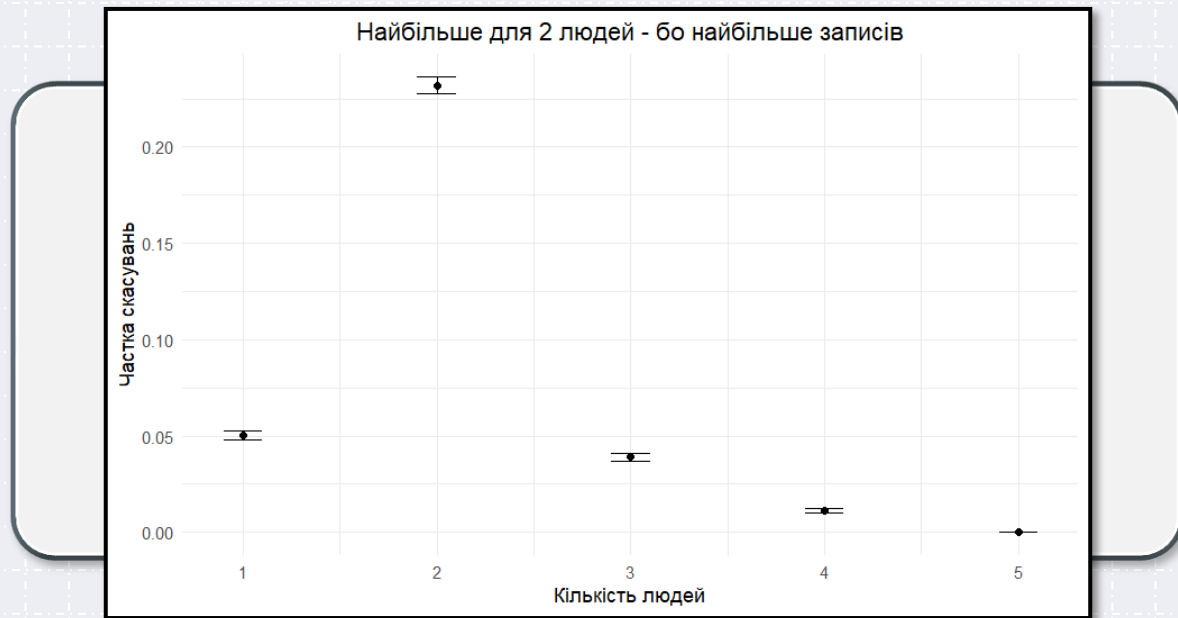
Обчислимо довірчі інтервали для часток скасованих записів, що приходить на людей, вказаних у бронюванні, розбивши на окремі графіки дорослих та дітей та дітей



# 03

Які характерні риси скасованих записів?

Обчислимо довірчі інтервали для скасованих записів, що  
приходиться кількість людей, вказаних у бронюванні



# 03

## Які характерні риси скасованих записів?

Гіпотеза Н<sub>0</sub>: невідмінені бронювання в середньому мають більше людей, ніж відмінені

```
Mean (Requests): 2.033594
Mean (No Requests): 1.917721
T-statistic: -16.15134
P-value: 5.556227e-59
95% Confidence Interval: -0.1299349 -0.1018124
```

Welch Two Sample t-test

```
data: no_of_people by booking_status_binary
t = -16.151, df = 23890, p-value < 2.2e-16
alternative hypothesis: true difference in means between group 0 and group 1 is less
than 0
95 percent confidence interval:
-Inf -0.1040726
sample estimates:
mean in group 0 mean in group 1
1.917721          2.033594
```

| mean<br><dbl> | sd<br><dbl> | n<br><int> | a<br><dbl> | b<br><dbl> |
|---------------|-------------|------------|------------|------------|
| 1.917721      | 0.6416413   | 23797      | 1.909568   | 1.925873   |
| mean<br><dbl> | sd<br><dbl> | n<br><int> | a<br><dbl> | b<br><dbl> |
| 2.033594      | 0.6370452   | 11877      | 2.022137   | 2.045051   |

# 03

## Які характерні риси скасованих записів?

Гіпотеза  $H_0$ : в середньому на не скасованих бронюваннях час до прибуття є більшим, ніж у скасованих

```
Mean (Requests): 139.1415
Mean (No Requests): 59.81611
T-statistic: -79.48999
P-value: 0
95% Confidence Interval: -81.28133 -77.36952
```

Welch Two Sample t-test

```
data: lead_time by booking_status_binary
t = -79.49, df = 16992, p-value < 2.2e-16
alternative hypothesis: true difference in means between group 0 and group 1 is less
than 0
95 percent confidence interval:
 -Inf -77.68388
sample estimates:
mean in group 0 mean in group 1
 59.81611      139.14153
```

| mean<br><dbl> | sd<br><dbl> | n<br><int> | a<br><dbl> | b<br><dbl> |
|---------------|-------------|------------|------------|------------|
| 59.81611      | 64.02325    | 23797      | 59.00267   | 60.62955   |
| mean<br><dbl> | sd<br><dbl> | n<br><int> | a<br><dbl> | b<br><dbl> |
| 139.1415      | 98.90445    | 11877      | 137.3628   | 140.9203   |

# 03

## Додатково

Які характерні риси скасованих записів?

Максимальне значення `lead_time`:

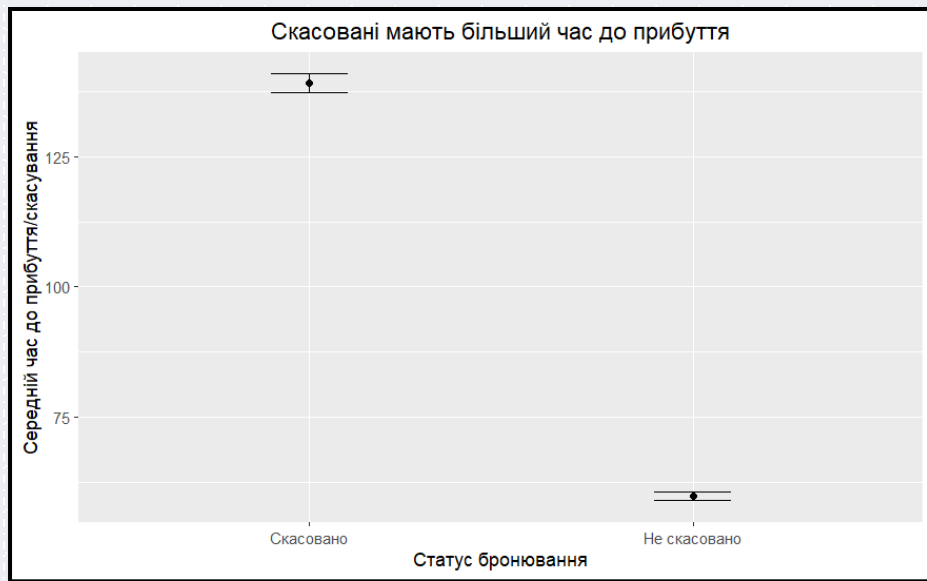
```
```{r}
max(hotel$lead_time)
```

[1] 443
```

Кількість відмінених і невідмінених записів за умови  $350 \leq \text{lead\_time} \leq 443$ :

| A tibble: 2 × 2          |                |
|--------------------------|----------------|
| booking_status<br><fctr> | count<br><int> |
| Canceled                 | 244            |
| Not_Canceled             | 24             |

Розглянемо отриманий результат графічно:



| mean     | sd       | n     | a        | b        |
|----------|----------|-------|----------|----------|
| <dbl>    | <dbl>    | <int> | <dbl>    | <dbl>    |
| 59.81611 | 64.02325 | 23797 | 59.00267 | 60.62955 |
| mean     | sd       | n     | a        | b        |
| <dbl>    | <dbl>    | <int> | <dbl>    | <dbl>    |
| 139.1415 | 98.90445 | 11877 | 137.3628 | 140.9203 |

# 03

## Які характерні риси скасованих записів?

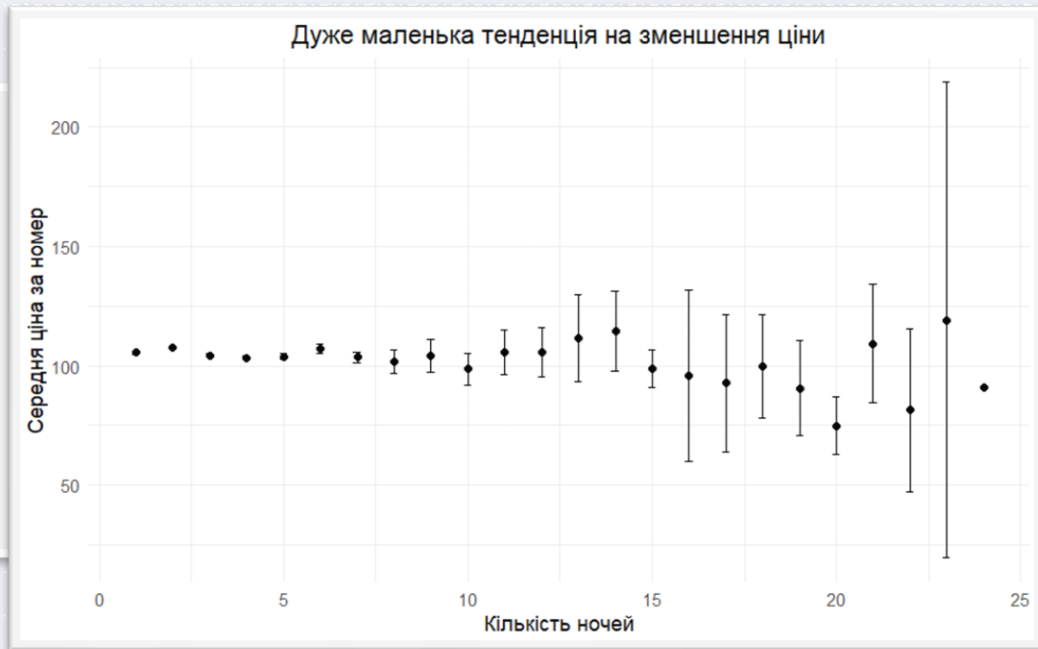
**Маємо проміжні висновки:**

- найбільша кількість скасування приходить на записи з двома людьми. В більшості випадків це саме двоє дорослих, що планували заселятися без дітей
- відмінені і невідмінені бронювання незначним чином відрізняються одне від одного по середній кількості людей на бронювання
- з неочевидних причин середня кількість часу від моменту бронювання до скасування приблизно вдвічі перевищує час до прибуття у нескасованих записів

# 04

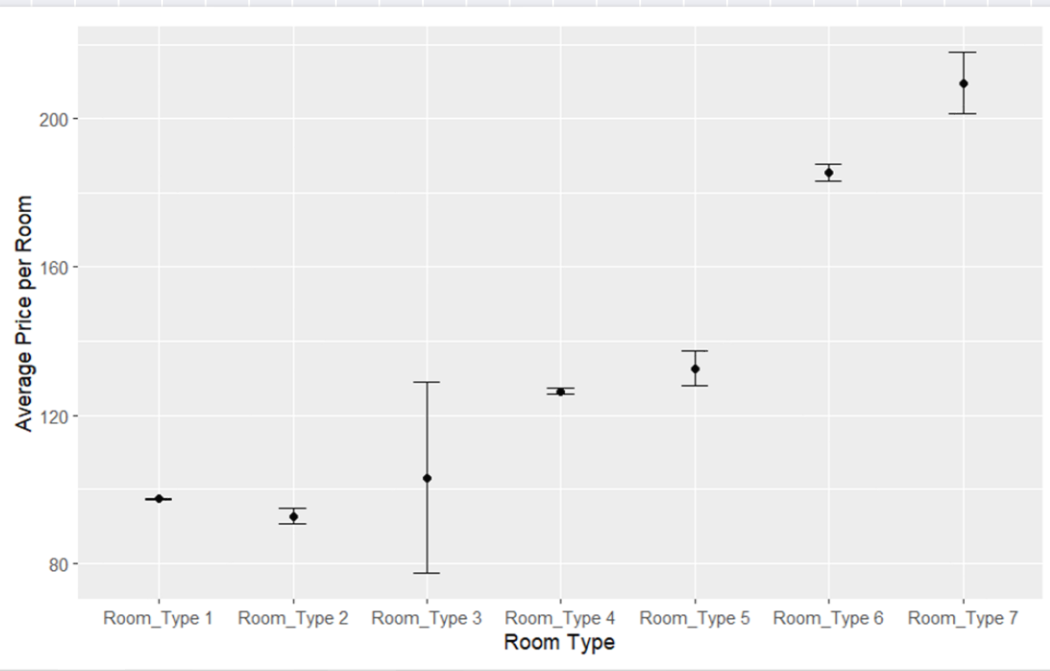
## Що впливає на кількість проведених ночей у готелі?

Побудуємо довірчі інтервали для середніх значень цін за кімнати по кожному значенню проведених ночей у готелі





## Додатково (1)



## 04

### Додатково (2)

Межі для середньої ціни кожного з типів кімнат:

A tibble: 7 × 3

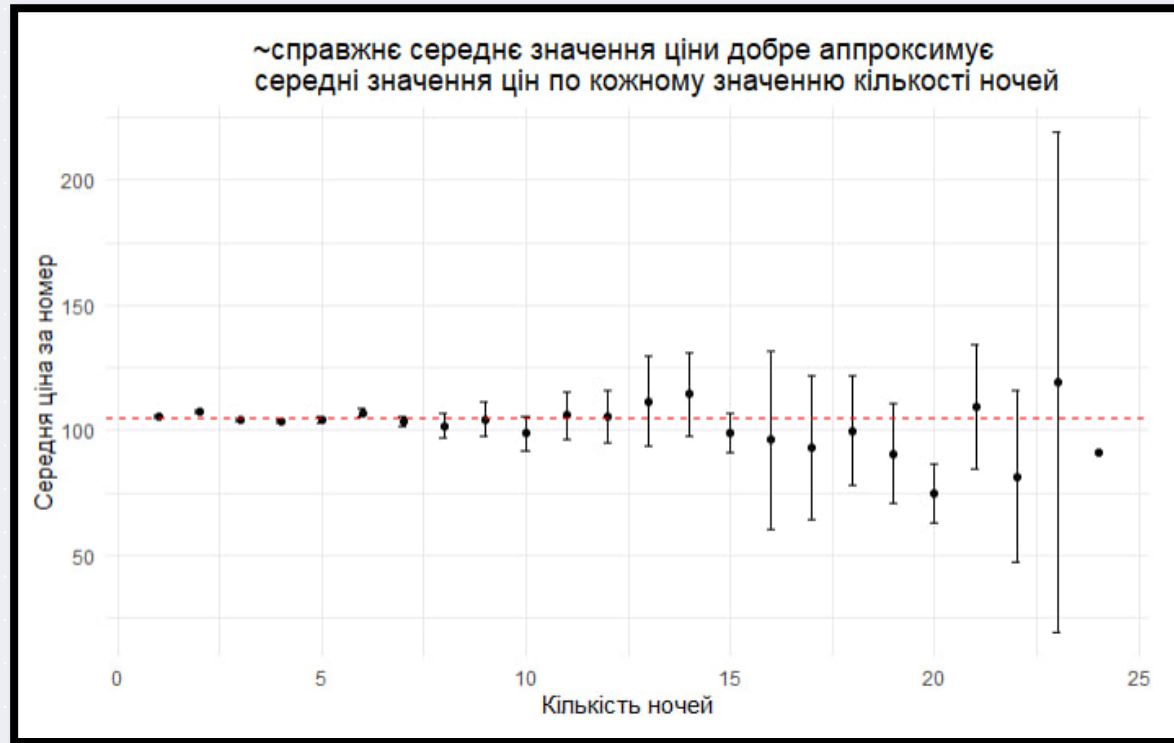
| room_type_reserved<br><fctr> | min_price<br><dbl> | max_price<br><dbl> |
|------------------------------|--------------------|--------------------|
| Room_Type 1                  | 12.0               | 365.00             |
| Room_Type 2                  | 12.0               | 284.10             |
| Room_Type 3                  | 65.0               | 130.00             |
| Room_Type 4                  | 12.0               | 375.50             |
| Room_Type 5                  | 12.0               | 250.00             |
| Room_Type 6                  | 18.6               | 349.63             |
| Room_Type 7                  | 52.0               | 306.00             |

7 rows

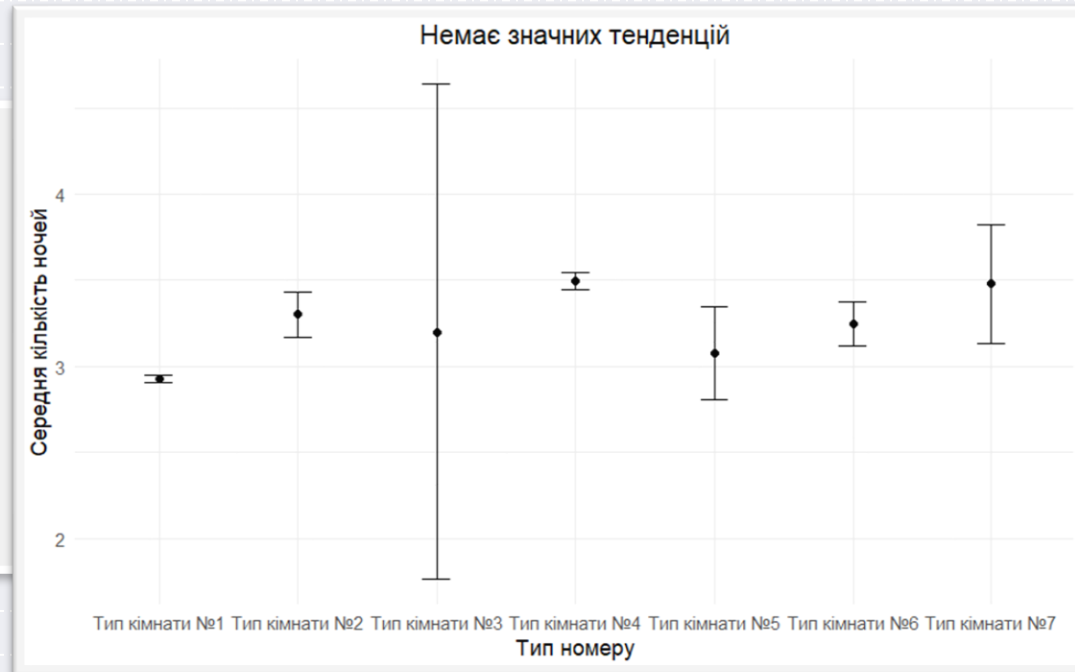
Довірчий інтервал для **avg\_price\_per\_room**:

| mean<br><dbl> | sd<br><dbl> | n<br><int> | a<br><dbl> | b<br><dbl> |
|---------------|-------------|------------|------------|------------|
| 105.1521      | 32.72643    | 35674      | 104.8125   | 105.4917   |

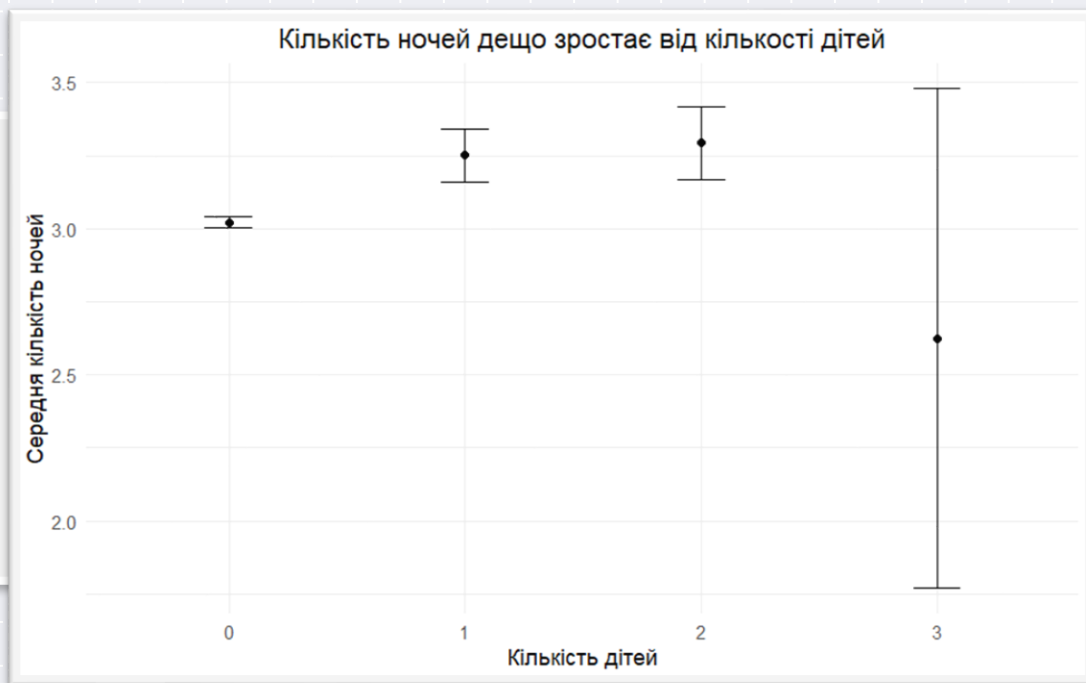
## Додатково (3)



Побудуємо довірчі інтервали для середніх значень кількості ночей по кожному типу кімнати:



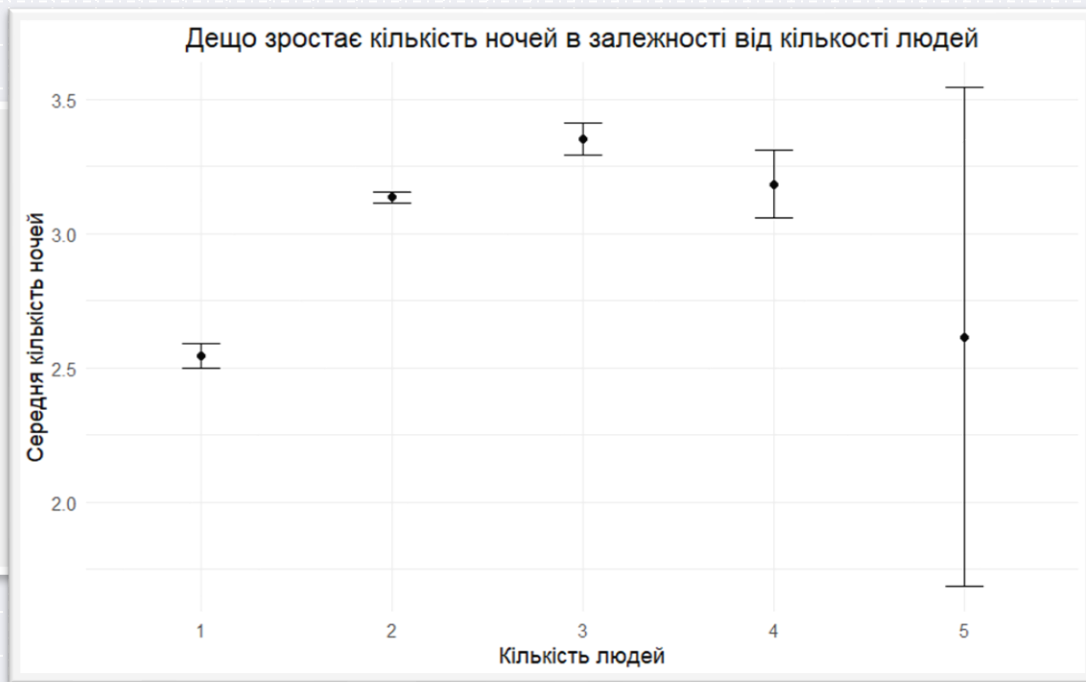
Побудуємо довірчі інтервали для  
кількості ночей залежно від кількості дітей



# 04

Що впливає на кількість проведених ночей у готелі?

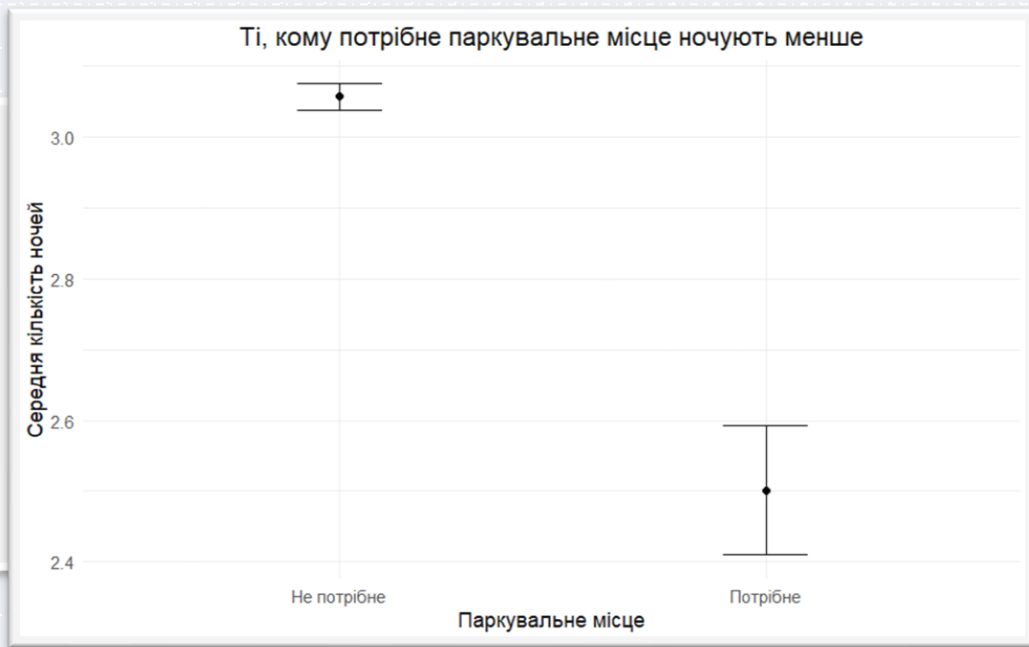
Побудуємо довірчі інтервали для  
кількості ночей залежно від кількості людей



# 04

Що впливає на кількість проведених ночей у готелі?

Побудуємо довірчі інтервали для  
кількості ночей залежно від паркувального місця



Побудуємо довірчі інтервали для  
кількості ночей залежно від часу до прибуття





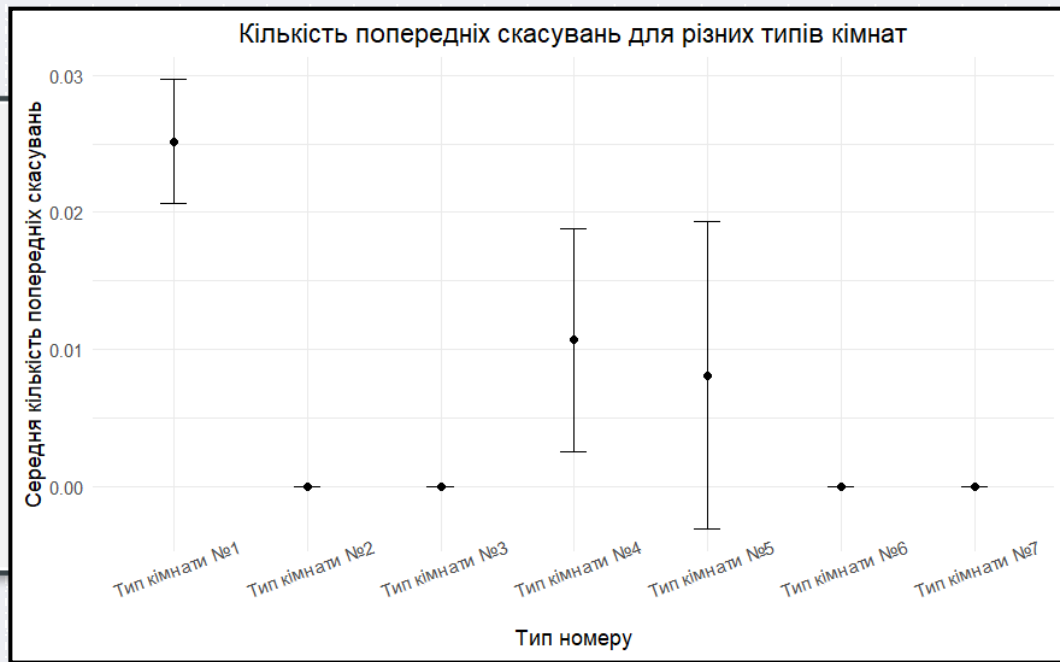
## 04

## Висновки

- незалежно від типу номеру, середня кількість заброньованих ночей здебільшого однакова
- наявність та кількість дітей практично не впливає на середню кількість ночей
- на найменшу в середньому кількість ночей лишається одна особа
- відвідувачі які приїзять власним автомобілем в середньому лишаються на помітно меншу кількість ночей
- середня кількість ночей зростає зі зростанням часу до прибуття десь до 170-200 годин, після чого закономірно спадає

## Що впливає на кількість попередніх скасування/не скасування?

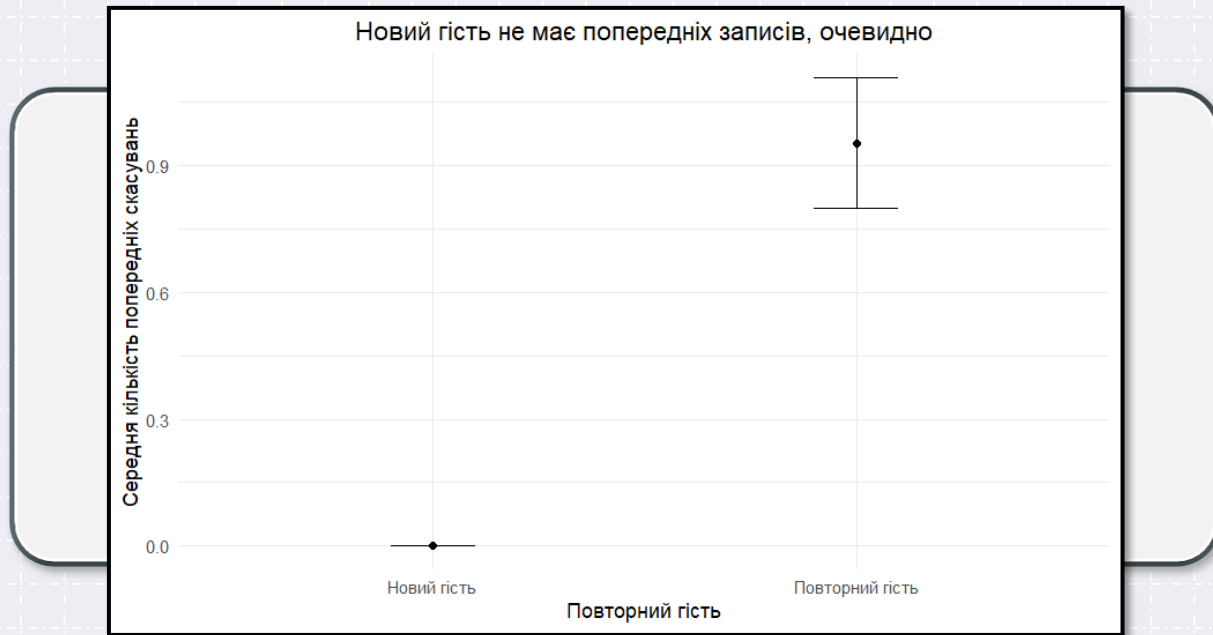
Можемо побудувати довірчі інтервали для середньої кількості попередніх скасування для кожного типу кімнати



# 05

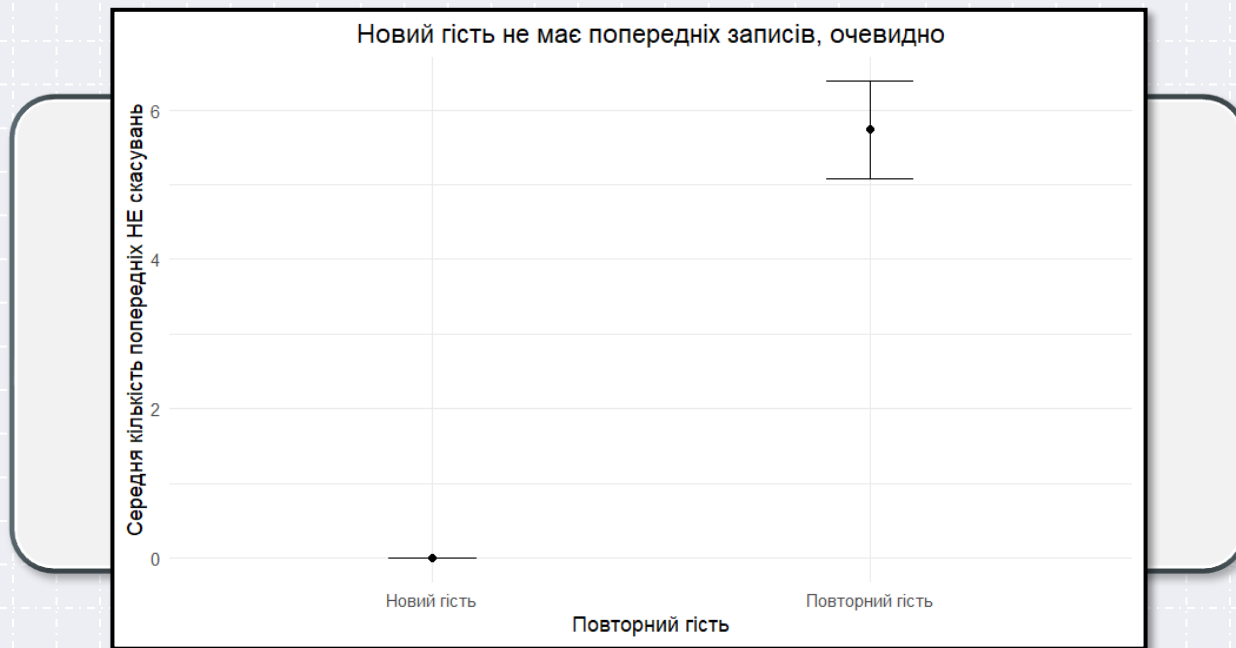
Що впливає на кількість попередніх скасування/не скасування?

Спробуємо побудувати схожі довірчі інтервали для кількості попередніх скасування, але згрупуємо дані за фактом повторного гостя



Слайд  
№43

Побудуємо аналогічний інтервал для не відмінених записів



## Формулювання гіпотез: Відмінені і невідмінені бронювання мають однакову кількість попередніх скасування бронювань

Welch Two Sample t-test

```
data: no_of_previous_cancellations by booking_status_binary
t = 6.853, df = 34714, p-value = 7.354e-12
alternative hypothesis: true difference in means between group 0 and group 1 is not
equal to 0
95 percent confidence interval:
 0.01694468 0.03052022
sample estimates:
mean in group 0 mean in group 1
 0.029289406      0.005556959

Mean (Canceled): 0.005556959
Mean (Not Canceled): 0.02928941
T-statistic: 6.852965
P-value: 7.233456e-12
95% Confidence Interval: 0.01694491 0.03051998
```

| mean<br><dbl> | sd<br><dbl> | n<br><int> | a<br><dbl>  | b<br><dbl>  |
|---------------|-------------|------------|-------------|-------------|
| 0.02928941    | 0.409922    | 23797      | 0.0240812   | 0.03449761  |
| mean<br><dbl> | sd<br><dbl> | n<br><int> | a<br><dbl>  | b<br><dbl>  |
| 0.005556959   | 0.2420221   | 11877      | 0.001204349 | 0.009909569 |

Протестуємо аналогічну гіпотезу:  
Відмінені і невідмінені бронювання мають  
однакову кількість попередніх НЕ скасуваннях бронювань

```
Welch Two Sample t-test

data: no_of_previous_bookings_not_canceled by booking_status_binary
t = 14.531, df = 24193, p-value < 2.2e-16
alternative hypothesis: true difference in means between group 0 and group 1 is not
equal to 0
95 percent confidence interval:
 0.1642063 0.2154134
sample estimates:
mean in group 0 mean in group 1
 0.191998991      0.002189105

Mean (Canceled): 0.002189105
Mean (Not Canceled): 0.191999
T-statistic: 14.53076
P-value: 7.735575e-48
95% Confidence Interval: 0.1642076 0.2154122
```

| mean<br><dbl> | sd<br><dbl> | n<br><int> | a<br><dbl>    | b<br><dbl>  |
|---------------|-------------|------------|---------------|-------------|
| 0.191999      | 2.006689    | 23797      | 0.1665033     | 0.2174947   |
| mean<br><dbl> | sd<br><dbl> | n<br><int> | a<br><dbl>    | b<br><dbl>  |
| 0.002189105   | 0.1297532   | 11877      | -0.0001444219 | 0.004522632 |

# 05

## Що впливає на кількість попередніх скасуваннях/нескасуваннях?

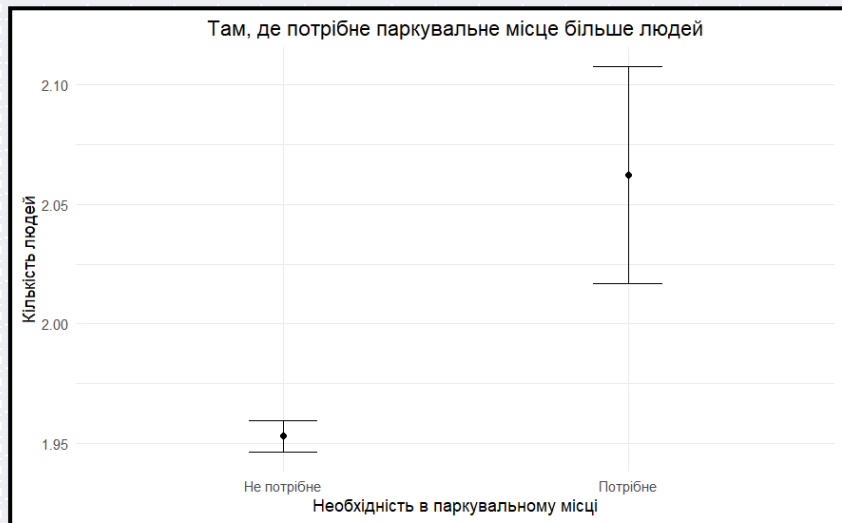
Маємо проміжні висновки:

Дослідження даного питання дало дуже слабкі результати, оскільки складно оцінити вплив на скасуваннях/нескасуваннях бронювань через дуже малу кількість записів з наявними даними про це. До того ж, у нас немає інформації про попередні записи зі скасуваннями/нескасуваннями, тому ми не можемо робити висновки щодо впливу. Отже дане дослідження не дало відповіді на дане питання, але і подальше його вивчення навряд чи принесе які-небудь результати через відсутність інформації про попередні записи.

# 06

## Які характерні риси бронювань з потребою у паркувальному місці?

Побудуємо інтервал для кількості людей, згрупувавши дані за необхідністю в паркувальному місці



Вважаючи, що ми говоримо про кількість людей, різниця здається зовсім невеликою. Тому, існує необхідність протестувати відповідну гіпотезу про рівність цих груп.

Welch Two Sample t-test

```
data: no_of_people by required_car_parking_space
t = -4.656, df = 1141.3, p-value = 3.603e-06
alternative hypothesis: true difference in means between group 0 and group 1 is not
equal to 0
95 percent confidence interval:
 -0.15522792 -0.06318716
sample estimates:
mean in group 0 mean in group 1
 1.952950         2.062157

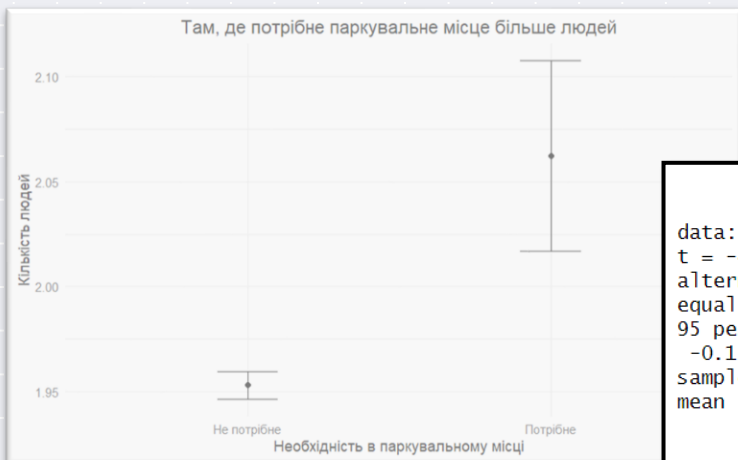
Mean (Needed): 2.062157
Mean (Not needed): 1.95295
T-statistic: -4.655982
P-value: 3.224399e-06
95% Confidence Interval: -0.1551791 -0.06323596
```



# 06

## Які характерні риси бронювань з потребою у паркувальному місці?

Побудуємо інтервал для кількості людей, згрупувавши дані за необхідністю в паркувальному місці



Вважаючи, що ми говоримо про кількість людей, різниця здається зовсім невеликою. Тому, існує необхідність протестувати відповідну гіпотезу про рівність цих груп

### Welch Two Sample t-test

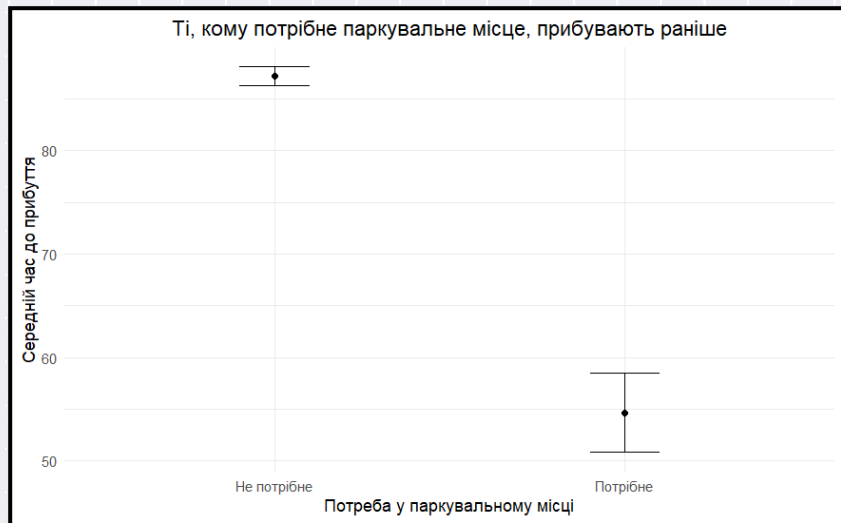
```
data: no_of_people by required_car_parking_space
t = -4.656, df = 1141.3, p-value = 3.603e-06
alternative hypothesis: true difference in means between group 0 and group 1 is not
equal to 0
95 percent confidence interval:
 -0.15522792 -0.06318716
sample estimates:
mean in group 0 mean in group 1
 1.952950      2.062157

Mean (Needed): 2.062157
Mean (Not needed): 1.95295
T-statistic: -4.655982
P-value: 3.224399e-06
95% Confidence Interval: -0.1551791 -0.06323596
```

# 06

## Які характерні риси бронювань з потребою у паркувальному місці?

Побудуємо інтервал для часу до прибуття, знову розділивши по групам за необхідністю в паркувальному місці



Люди, яким не потрібно паркувальне місце в середньому платять більше за тих, кому паркувальне місце потрібне

Welch Two Sample t-test

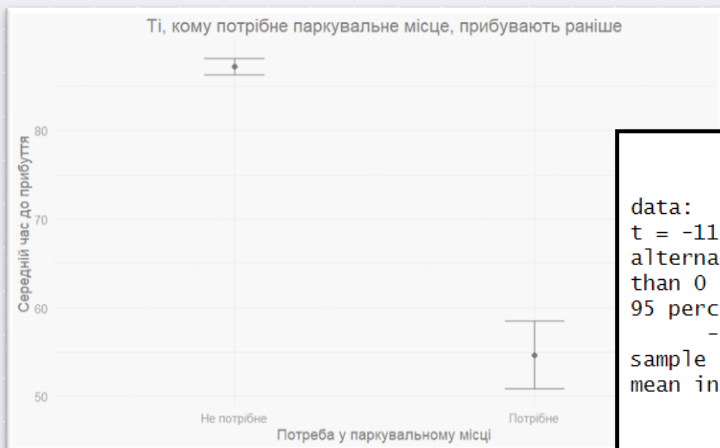
```
data: avg_price_per_room by required_car_parking_space
t = -11.344, df = 1138.6, p-value < 2.2e-16
alternative hypothesis: true difference in means between group 0 and group 1 is less than 0
95 percent confidence interval:
 -Inf -11.86987
sample estimates:
mean in group 0 mean in group 1
 104.7263      118.6111

Mean (Needed): 118.6111
Mean (Not needed): 104.7263
T-statistic: -11.34357
P-value: 3.991131e-30
95% Confidence Interval: -16.28391 -11.4858
```

# 06

## Які характерні риси бронювань з потребою у паркувальному місці?

Побудуємо інтервал для часу до прибуття, знову розділивши по групах за необхідністю в паркувальному місці



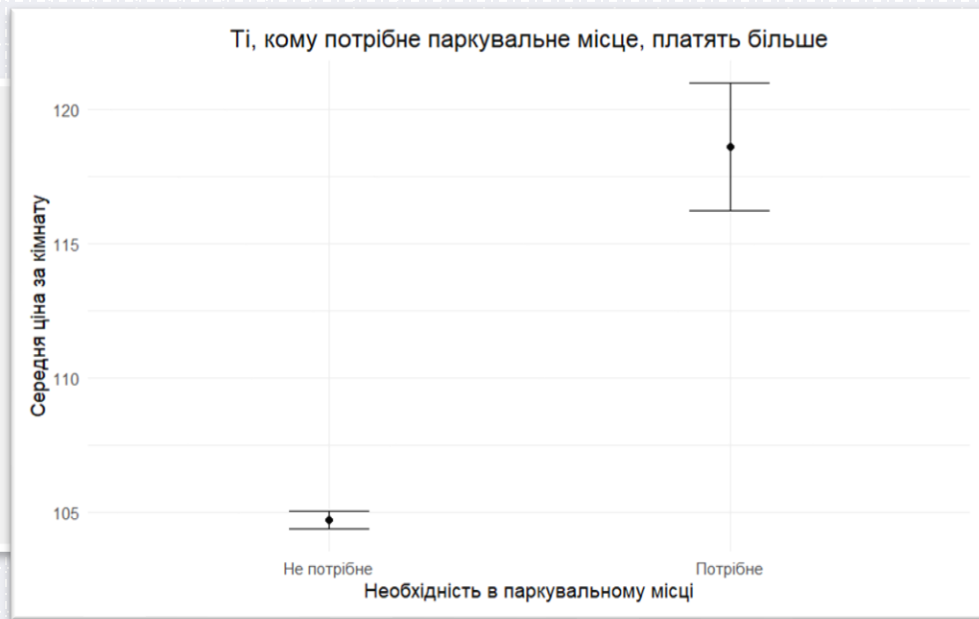
**Люди, яким не потрібно паркувальне місце в середньому платять більше за тих, кому паркувальне місце потрібне**

### Welch Two Sample t-test

```
data: avg_price_per_room by required_car_parking_space
t = -11.344, df = 1138.6, p-value < 2.2e-16
alternative hypothesis: true difference in means between group 0 and group 1 is less than 0
95 percent confidence interval:
    -Inf -11.86987
sample estimates:
mean in group 0 mean in group 1
    104.7263      118.6111
```

Mean (Needed): 118.6111  
Mean (Not needed): 104.7263  
T-statistic: -11.34357  
P-value: 3.991131e-30  
95% Confidence Interval: -16.28391 -11.4858

Побудуємо відповідний довірчий інтервал для середньої ціни за кімнату в залежності від необхідності в паркуванні



# 06

## Які характерні риси бронювань з потребою у паркувальному місці?

Маємо проміжні висновки:

Детальніше досліджуючи відмінність для групи клієнтів, що потребують паркувальне місце, було виявлено декілька цікавих спостережень: в даних записах більше осіб, час до прибуття менший і платять за номер вони дещо більше.

До перших двох спостережень немає ніяких запитань - достатньо природньо, що на машині можна привезти більше людей (більше ніж 2 в записах без паркувального місця), або взяти з собою всю сім'ю; можна набагато швидше дістатися готелю, тому і час до прибуття менший.

А от з ціною за кімнату вже цікавіше - дана група людей платить більше, бо має на це гроші (бо мають автомобіль), чи є інші фактори які на це впливають? Це питання потребує подальшого вивчення

## Які нові цікаві відомості про повторних гостей?

Гіпотеза  $H_0$ : загальна кількість гостей, які приїжджають вперше і не вперше – однакова

Welch Two Sample t-test

```
data: no_of_people by repeated_guest
t = 48.099, df = 879.91, p-value < 2.2e-16
alternative hypothesis: true difference in means between group 0 and group 1 is not
equal to 0
95 percent confidence interval:
 0.7288707 0.7908840
sample estimates:
mean in group 0 mean in group 1
 1.973360      1.213483

Mean (Requests): 1.213483
Mean (No Requests): 1.97336
T-statistic: 48.09885
P-value: 0
95% Confidence Interval: 0.7289133 0.7908413
```

| mean<br><dbl> | sd<br><dbl> | n<br><int> | a<br><dbl> | b<br><dbl> |
|---------------|-------------|------------|------------|------------|
| 1.97336       | 0.6362855   | 34873      | 1.966682   | 1.980039   |
| mean<br><dbl> | sd<br><dbl> | n<br><int> | a<br><dbl> | b<br><dbl> |
| 1.213483      | 0.4365982   | 801        | 1.183248   | 1.243718   |

Так само протестуємо нульову гіпотезу про рівність (в середньому) оплати для гостей, що прибули вперше і не вперше

#### Welch Two Sample t-test

```
data: avg_price_per_room by repeated_guest
t = 41.497, df = 904.93, p-value < 2.2e-16
alternative hypothesis: true difference in means between group 0 and group 1 is not
equal to 0
95 percent confidence interval:
 28.27714 31.08463
sample estimates:
mean in group 0 mean in group 1
 105.81853          76.13764

Mean (Requests): 76.13764
Mean (No Requests): 105.8185
T-statistic: 41.49704
P-value: 0
95% Confidence Interval: 28.27901 31.08275
```

| mean<br><dbl> | sd<br><dbl> | n<br><int> | a<br><dbl> | b<br><dbl> |
|---------------|-------------|------------|------------|------------|
| 105.8185      | 32.66498    | 34873      | 105.4757   | 106.1614   |

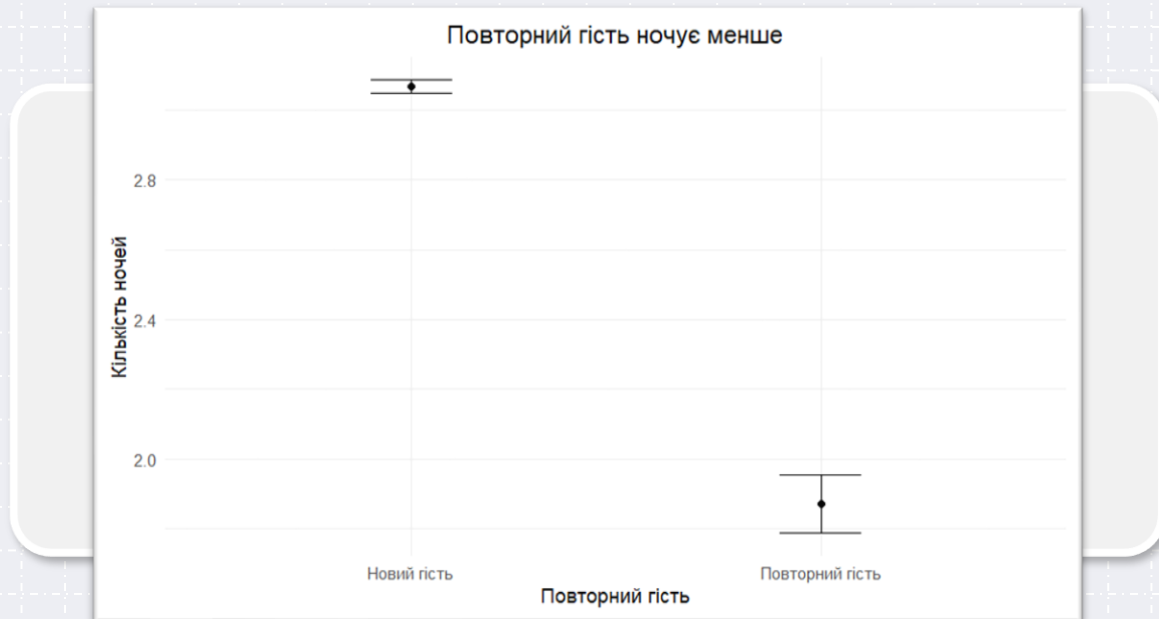
  

| mean<br><dbl> | sd<br><dbl> | n<br><int> | a<br><dbl> | b<br><dbl> |
|---------------|-------------|------------|------------|------------|
| 76.13764      | 19.62838    | 801        | 74.77834   | 77.49694   |

07

Які нові цікаві відомості про  
повторних гостей?

Побудуємо довірчий інтервал для середньої кількості ночей  
згрупувавши дані за повторністю гостя



Слайд  
№56



Спробуємо оцінити незалежність записів з повторним гостем від статусу бронювання і необхідності в паркувальному місці

```
      booking_status
repeated_guest Canceled Not_Canceled
0      11861      23012
1         16         785
```

Pearson's Chi-squared test

```
data: cont_tab
X-squared = 361.36, df = 1, p-value < 2.2e-16
```

```
      required_car_parking_space
repeated_guest      0      1
0      33913      960
1       667      134
```

Pearson's Chi-squared test

```
data: cont_tab
X-squared = 514.53, df = 1, p-value < 2.2e-16
```

Гіпотеза  $H_0$ : частка записів з менш ніж 2 людьми перевищує 21%

1-sample proportions test with continuity correction

```
data: observed_count out of nrow(hotel), null probability 0.21
X-squared = 11.426, df = 1, p-value = 0.0003622
alternative hypothesis: true p is less than 0.21
95 percent confidence interval:
 0.0000000 0.2062342
sample estimates:
      p
0.2026966
```

# 07

## Які нові цікаві відомості про повторних гостей?

Маємо проміжні висновки:

З проведеного дослідження щодо особливостей повторного гостя можна зробити висновки, що така група відрізняється за багатьма параметрами: в середньому їх приїжджає менше осіб, вони платять менше, ночують менше ночей. Нічого не можна сказати про незалежність даних спостережень, тому вони потребують подальшого вивчення

The slide features a light gray background with a white dashed grid. On the left and right edges, there are decorative elements consisting of horizontal bars of varying lengths. The bars on the left are primarily blue, with some gray bars interspersed. The bars on the right are primarily gray, with some blue bars interspersed. These bars appear to be of different heights and are arranged in a way that creates a sense of depth and structure.

**Дякую за увагу!**