

НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ УНІВЕРСИТЕТ УКРАЇНИ
«КИЇВСЬКИЙ ПОЛІТЕХНІЧНИЙ ІНСТИТУТ
ІМЕНІ ІГОРЯ СІКОРСЬКОГО»
Факультет прикладної математики
Кафедра прикладної математики

Звіт
з лабораторної роботи № 2: «Статистичне виведення»
із дисципліни «Аналіз даних»

Виконали:

Гармаш О. Є.

Хок М. Ш.

Куцалаба Н. В.

Маслов Н.Р.

Керівник:

ст. викладач Тавров Д. Ю.

1 Мотивація дослідження

1.1 Вступ

Літаки завжди були одним із перших варіантів для подорожей через їхню зручність і безпеку. З постійним підвищенням рівня життя людей, зростають групи клієнтів цивільної авіації, і люди висувають більш високі вимоги до якості авіаційних послуг. Прогнозування задоволеності пасажирів літака та визначення основних факторів, що на неї впливають, можуть допомогти авіакомпаніям покращити свої послуги та отримати переваги в складних ситуаціях і конкуренції. Таким чином, авіакомпанії повинні своєчасно досліджувати задоволеність пасажирів різними послугами та загальну задоволеність, щоб точно розуміти якість обслуговування існуючих послуг. Крім того, авіакомпанії повинні чітко розуміти основні фактори, що впливають на задоволеність пасажирів, і сформулювати відповідні стратегії для покращення якості обслуговування, щоб максимізувати загальну задоволеність пасажирів авіакомпанією та підвищити лояльність пасажирів.

Виходячи з вищезазначених проблем, у цьому дослідженні в якості об'єкта дослідження використовується повна інформація про пасажирів і результати опитування щодо задоволеності окремими факторами рейсу.

1.2 Дослідницькі питання та висновки з ЛР1

В даному дослідженні була спроба підтвердити або спростувати деякі гіпотези, які сформувалися в процесі EDA, провести статистичне виведення, щоб дізнатися варіативність усіх статистик, які ми обчислили в ЛР1 та, можливо, отримати нові гіпотези. Фактично в даній лабораторній ми ставимо під сумнів висновки, які впливають в результаті першої лабораторної на основі обрахованих статистик. Гіпотези першої лабораторної та відповідні висновки:

- **Гіпотеза:** *Існують вікові категорії що мають меншу частку задоволеності рейсом за інші*

Висновок ЛР1: *На відміну від людей середнього віку, діти та люди похилого віку здебільшого незадоволені рейсами авіакомпанії*

- **Гіпотеза:** *Існують вікові категорії, для яких вплив деяких факторів задоволеності менший за інші.*

Висновок ЛР1: *Існує проблема в комфортності сидінь в літаках для дітей. Низька задоволеність онлайн-реєстрацією серед людей похилого віку. : Комфортабельністю можна нехтувати при невеликих дистанціях рейсів.*

- **Гіпотеза:** *Пасажири, що подорожували в персональних цілях залишаються менш задоволеними перельотами.*

Висновок ЛР1: *Націленість компанії на бізнес мандрівки*

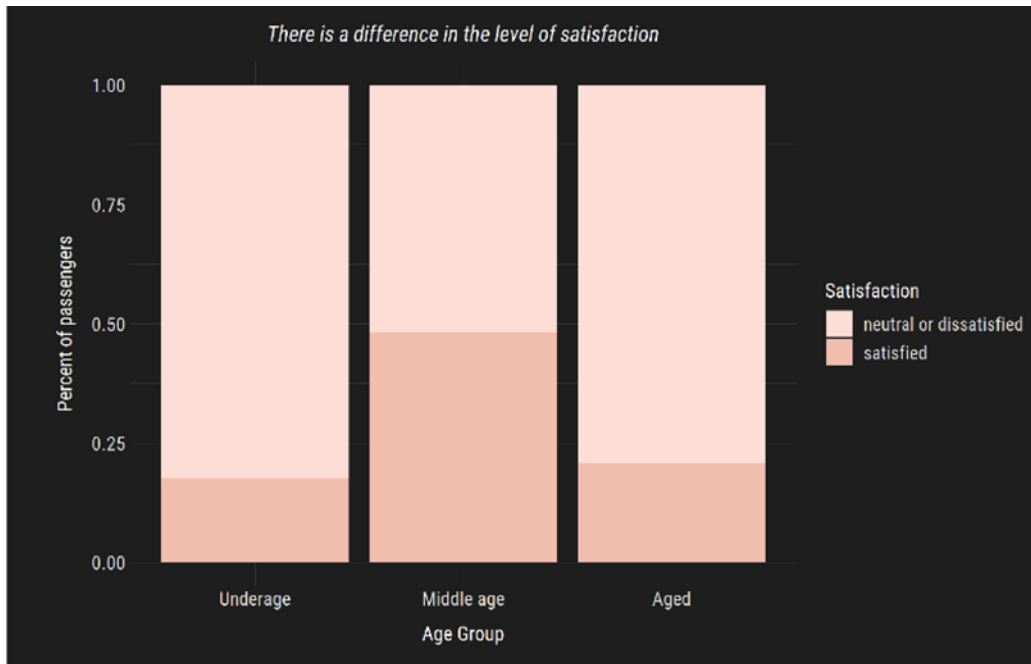
- **Гіпотеза:** *Існують фактори, які по-різному впливають на людей, подорожуючих різними класами.*

Висновок ЛР1: *Найбільш впливовіші фактори – online boarding, inflight entertainment, inflight wifi service*

2 Статистичне виведення

2.1 Перша група статистик

В даному розділі буде інформація пов'язана з аналізом групи статистик, що мають асимптотичний нормальний розподіл, дисперсію якого можна просто оцінити.



З даного графіка починається EDA в минулій ЛР і ми висували **гіпотезу 1**: "Існують вікові категорії що мають меншу частку задоволеності рейсом за інші." Побудувавши довірчі інтервали ми провели двосторонні тести на рівні значущості 0.05 на гіпотезах про рівність часток спостережень відповідно для кожної пари. Маємо наступні гіпотези:

H0: Статистично значущої різниці між частками спостережень немає.

H1: Існує статистично значуща різниця між частками спостережень.

Результати:

"Two side test between Underage and Middle age"

X-squared = 3780 **p-value < 2.2e-16**

95 percent confidence interval:

-0.3132165 -0.2977259

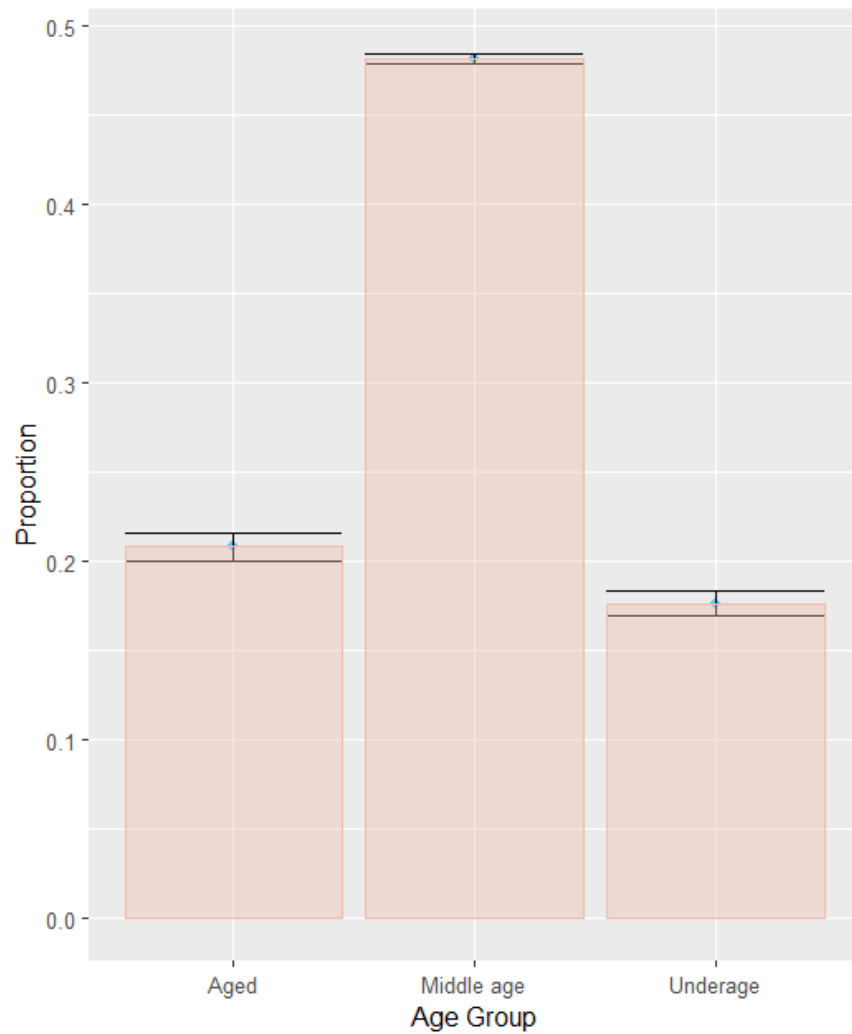
"Two side test between Underage and Aged"

X-squared = 34.563, **p-value = 4.128e-09**

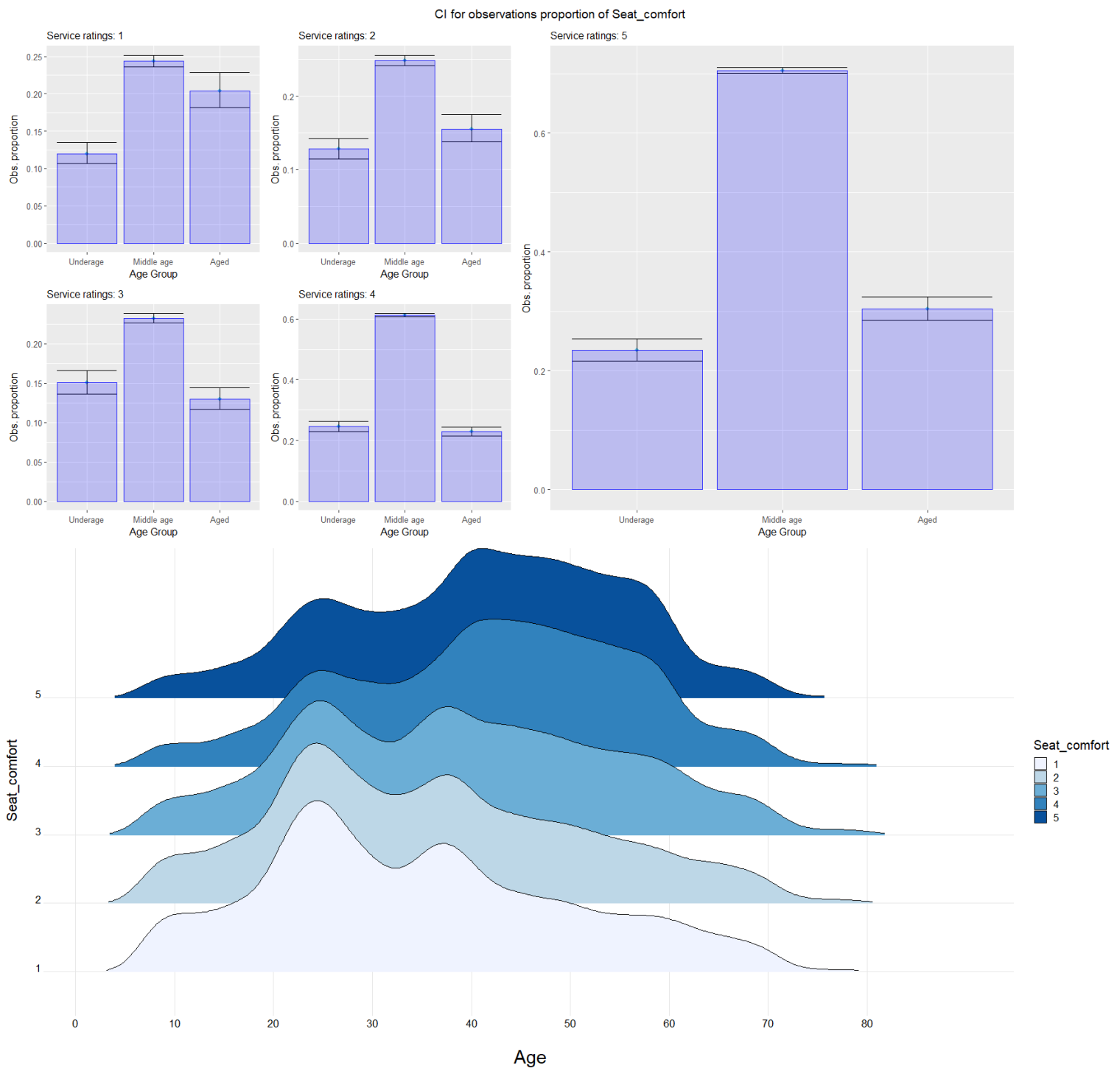
95 percent confidence interval:

-0.04270875 -0.02122165

"Two side test between Middle age and Aged"
X-squared = 2771.4, df = 1, p-value < 2.2e-16
95 percent confidence interval:
0.2649774 0.2820346



Далі ми нависили довірчи інтервали на всі статистики які знаходилися в другій гіпотезі ЛР1. А саме «Існують вікові категорії, для яких вплив деяких факторів задоволеності менший за інші».



На графіку вище ми розбили людей на групи – ті, хто поставив оцінку 1, 2, 3, 4, 5, а потім відповідно на вікові групи. В кожній такій групі ми тестували двосторонні гіпотези на рівність частки спостережень та перевіряли висновки зроблені в минулій лабораторній і помітили, що в деяких групах (наприклад в випадку з

Seat comfort), ми не можемо сказати, що люди з групи Underage поставили оцінку нижчу ніж Aged, бо $p_value = 0.1653$, що більше за 0.05 проте у всіх інших групах – висновок був підтверджений.

Результат двостороннього тесту між Underage та Aged серед тих хто поставив оцінку сервісу

Seat Comfort – 4.

Service rating: 4 Two sided prop.test between Underage and Aged"

X-squared = 1.9249, **p-value = 0.1653**

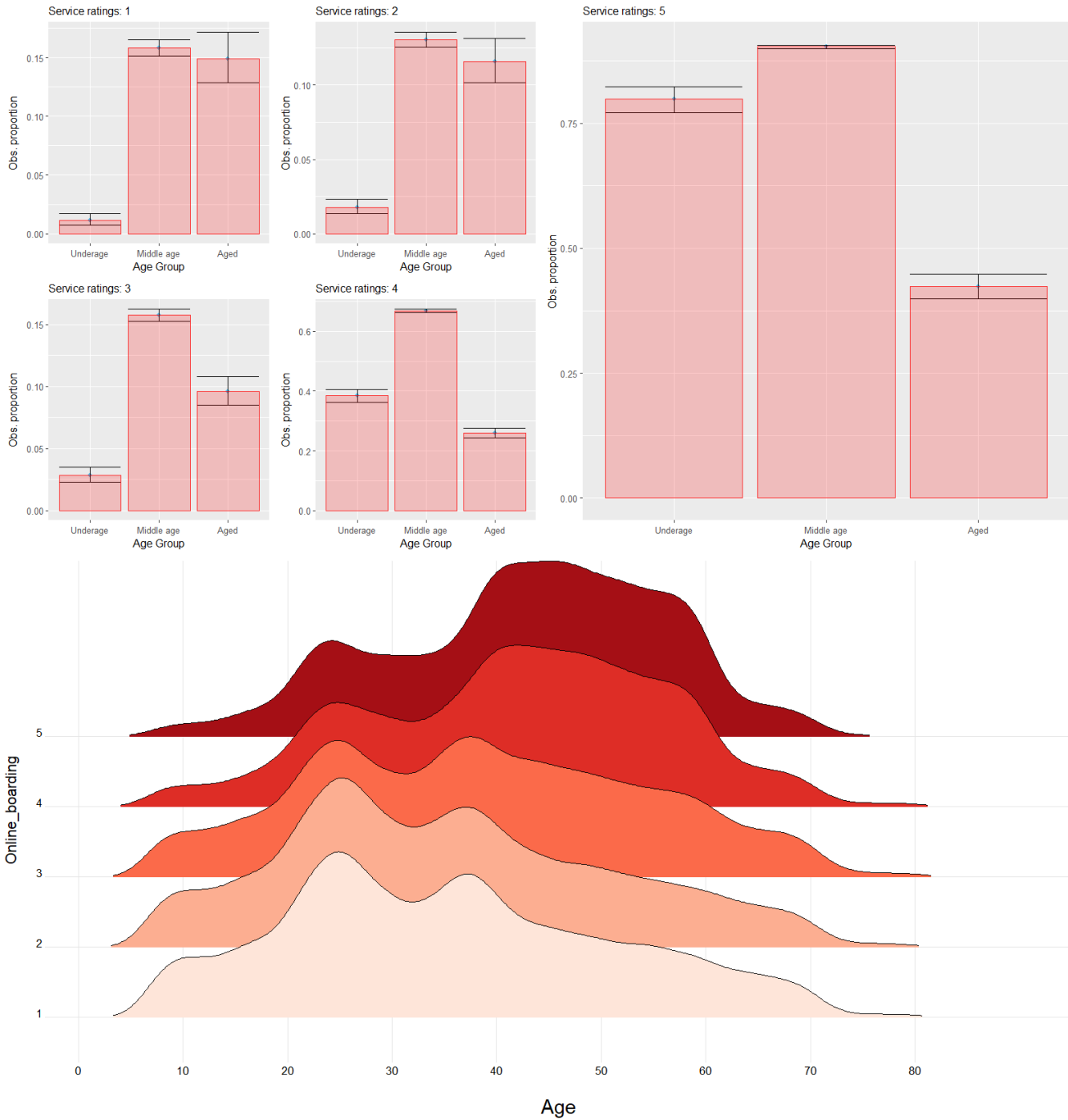
95 percent confidence interval:

-0.00666009 0.03935418

Всі наступні графіки демонструють ідентичний алгоритм, ми провели ті ж тести на предмет виявлення відмінностей від наших минулих висновків. Ті з них, що були спростовані будуть виведені під графіками як виноска, всі інші можна побачити лише в коді, тому що показувати результати 15 тестів по 7 графіках нечитабельно та немає сенсу.

Нище наведені графіки в яких гіпотеза підтверджується. Це може бути пов'язане з тим, що люди похилого віку можуть бути більш терпимі до незручностей або менш схильні до скарг. Вони можуть краще розуміти обмеження та складнощі, пов'язані з наданням послуг, і бути більш толерантними до них. Або мають більше життєвого досвіду, що може позитивно впливати на їхню задоволеність. Вони можуть мати більш розуміюче ставлення до різних ситуацій і легше оцінювати сервіс. Та й взагалі з віком змінюються пріоритети та очікування людей. Люди похилого віку можуть більше цінувати комфорт, зручність та спокій, що може призводити до більшої задоволеності від певних факторів, таких як комфорт сидінь у випадку Seat Comfort.

CI for observations proportion of Online_boarding



Гіпотеза 2

H0: Немає значущої різниці в частках задоволеності пасажирів з рейтингом 1, 2 сервісу Online Boarning між особами середнього віку та похилого віку.

H1: Існує значуща різниця в частках задоволеності пасажирів з рейтингом 1, 2 сервісу Online Boarning між особами середнього віку та похилого віку.

[1] "Service rating: 1 Two sided prop.test between Middle age and Aged"

X-squared = 0.55953, **p-value = 0.4545**

95 percent confidence interval:

-0.01371907 0.03225069

[1] "Service rating: 2 Two sided prop.test between Middle age and Aged"

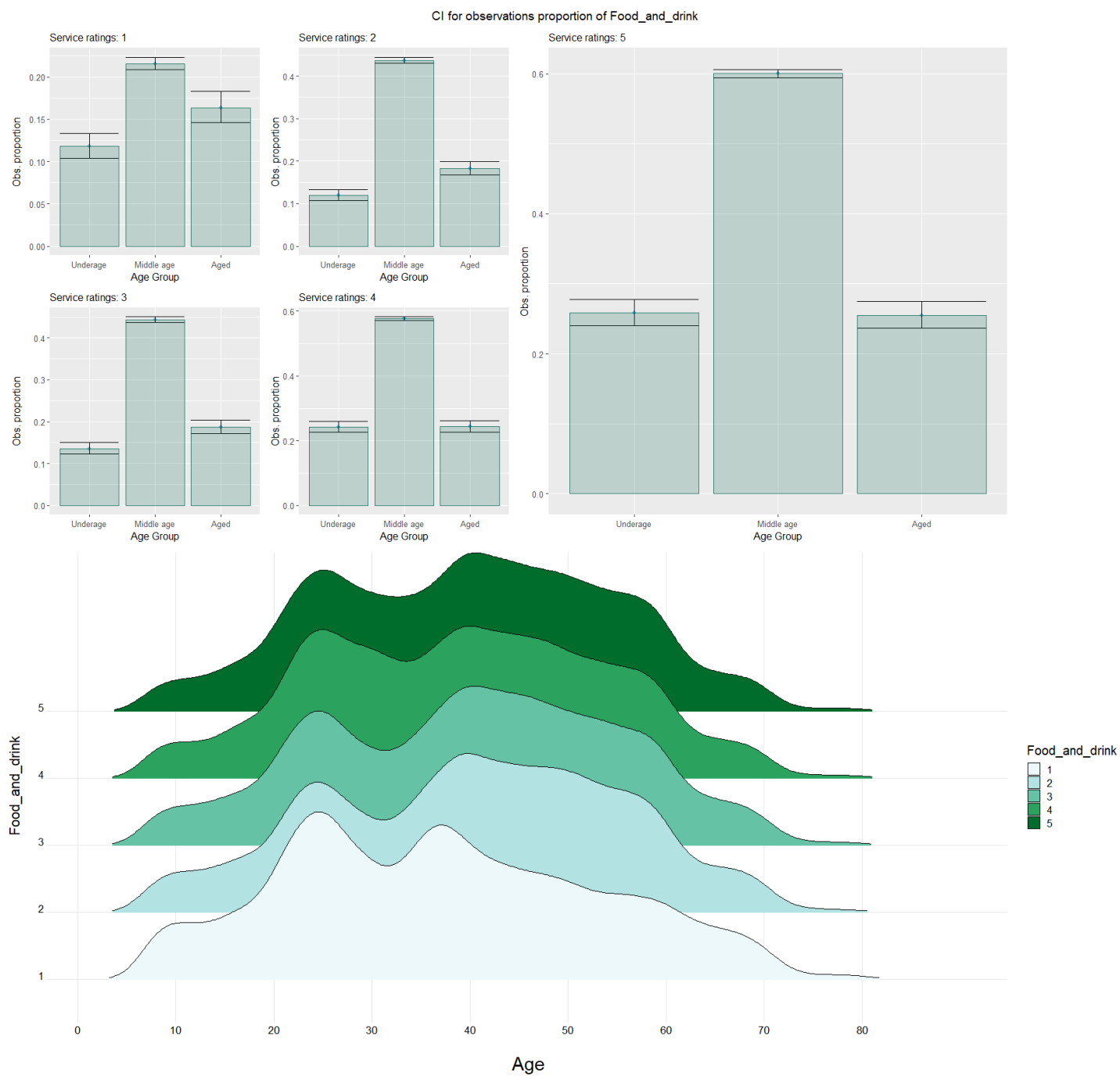
X-squared = 2.9166, **p-value = 0.08767**

95 percent confidence interval:

-0.001441608 0.030746398

У випадку "Service Online_boarding rating: 1" p-value = 0.4545, що більше за рівень значущості 0.05. Отже, немає достатніх доказів для відхилення нульової гіпотези H_0 . Ми не маємо підстав вважати, що середній рейтинг сервісу відрізняється значимо між особами середнього віку та похилого віку.

У випадку "Service Online_boarding rating: 2" p-value = 0.08767, що більше за рівень значущості 0.05, але близьке до нього. Ми не маємо достатніх доказів для відхилення нульової гіпотези H_0 , але є можливість наявності слабкої статистичної значущості. Це може вказувати на існування можливої різниці в середньому рейтингу сервісу між особами середнього віку та похилого віку, але потрібні додаткові дослідження для підтвердження цього.



Гіпотеза 3

H0: Люди похилого віку більш вимогливі, ніж діти.

H1: Люди похилого віку менш вимогливі, ніж діти.

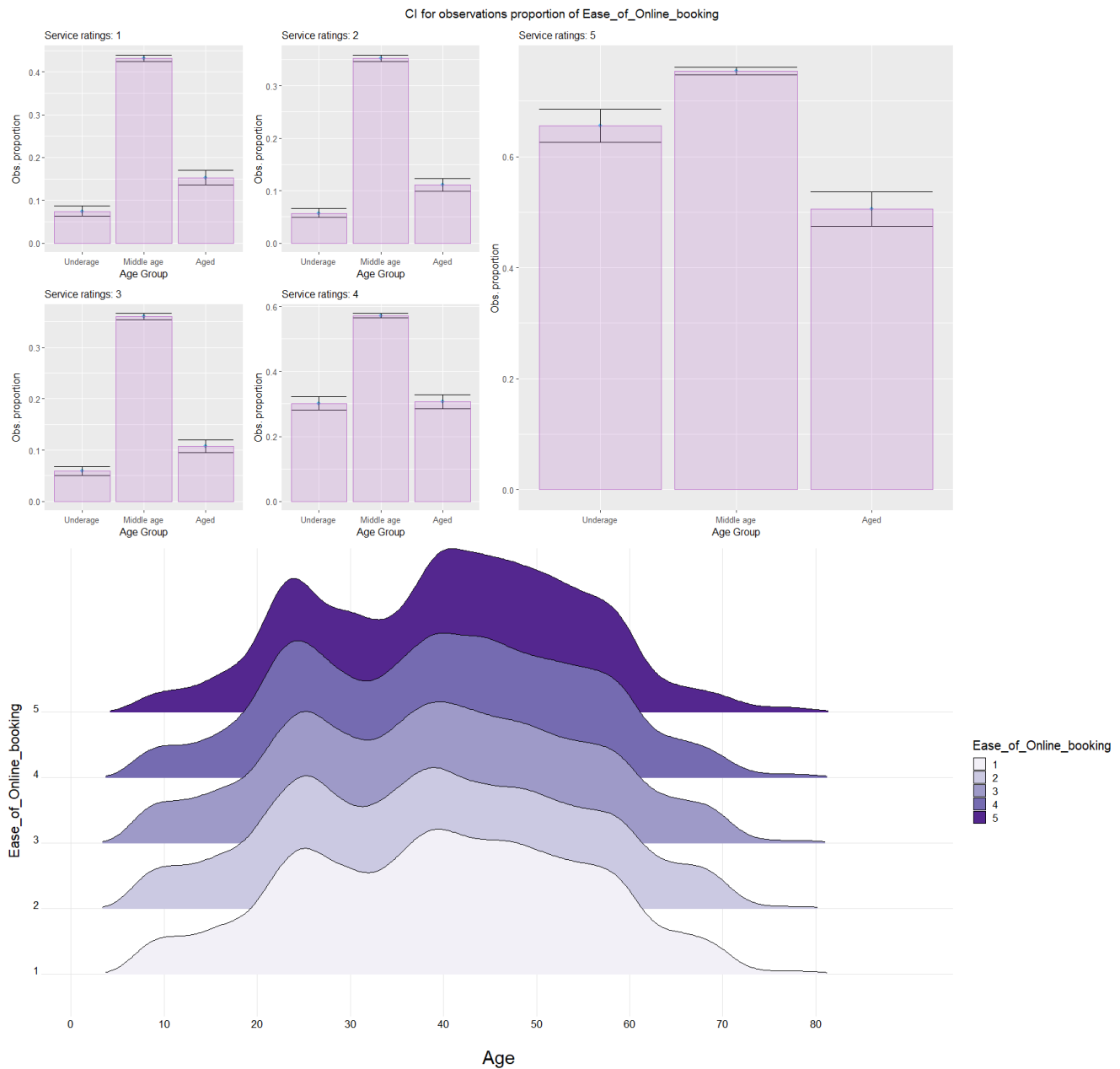
```
[1] "Service rating: 4 Two sided prop.test between Underage and Aged"
X-squared = 0.0039733, p-value = 0.9497
95 percent confidence interval:
-0.02643792 0.02397504
sample estimates:
prop 1 prop 2
0.2417983 0.2430297
```

```
[1] "Service rating: 5 Two sided prop.test between Underage and Aged"
X-squared = 0.047532, p-value = 0.8274
95 percent confidence interval:
-0.02404032 0.03105687
sample estimates:
prop 1 prop 2
0.2584600 0.2549518
```

У випадку "Service Food_and_drink rating: 4" $p\text{-value} = 0.9497$, що набагато більше за рівень значущості 0.05. Отже, немає достатніх доказів для відхилення нульової гіпотези H_0 . Ми не маємо підстав вважати, що люди похилого віку більш вимогливі, ніж діти, щодо рейтингу сервісу щодо їжі та напоїв.

У випадку "Service Food_and_drink rating: 5" $p\text{-value} = 0.8274$, що також набагато більше за рівень значущості 0.05. Знову ж таки, немає достатніх доказів для відхилення нульової гіпотези H_0 . Ми не маємо підстав вважати, що люди похилого віку більш вимогливі, ніж діти, щодо рейтингу сервісу щодо їжі та напоїв.

Отже, на підставі отриманих $p\text{-value}$ значень, ми не маємо достатніх доказів для підтримки альтернативної гіпотези H_1 . Наші дані не підтверджують, що люди похилого віку більш вимогливі, ніж діти, щодо рейтингу сервісу щодо їжі та напоїв.



Гіпотеза 4

H0: Дітям і людям похилого віку однаково зручно оформлювати онлайн-бронювання.

H1: Дітям зручніше оформлювати онлайн-бронювання, ніж людям похилого віку.

```
[1] "Service rating: 4 Two sided prop.test between Underage and Aged"
X-squared = 0.07401, p-value = 0.7856
95 percent confidence interval:
-0.03517774 0.02577029
sample estimates:
prop 1    prop 2 
0.3017751 0.3064789
```

У випадку "Service Ease_of_online_booking rating: 4" $p\text{-value} = 0.7856$, що набагато більше за рівень значущості 0.05. Отже, немає достатніх доказів для відхилення нульової гіпотези H_0 . Ми не маємо підстав вважати, що дітям зручніше оформлювати онлайн-бронювання, ніж людям похилого віку.

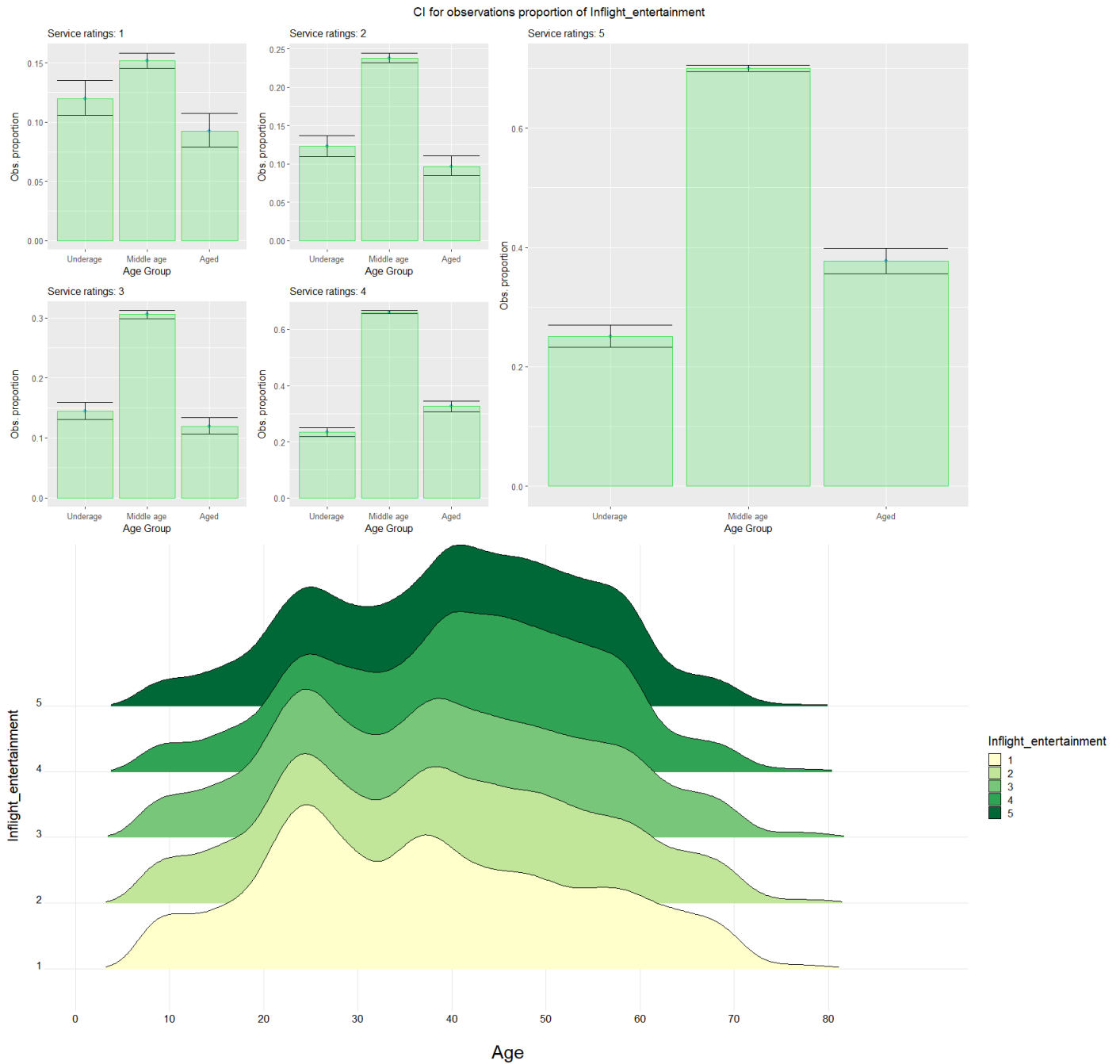
Отже, на підставі отриманого $p\text{-value}$ значення, ми не маємо достатніх доказів для підтримки альтернативної гіпотези H_1 . Наші дані не підтверджують, що дітям зручніше оформлювати онлайн-бронювання, ніж людям похилого віку.

Проміжний висновок

Майже всі альтернативні гіпотези справдилися крім однієї. (скрізь є різниця часток спостережень). На підставі результатів двосторонніх тестів між групами Underage та Aged для різних оцінок сервісу, можна зробити загальний висновок, що вік впливає на сприйняття сервісу, і зазвичай люди середнього віку (Middle Age) ставлять вищі оцінки, ніж інші вікові групи.

Зазначено, що "майже всі альтернативні гіпотези справдилися, крім однієї, де є різниця часток спостережень". Це означає, що для більшості оцінок сервісу (крім однієї), спостерігається статистично значуща різниця між групами в залежності від віку. Конкретно для оцінки, де середній вік отримав найвищу оцінку (Middle Age > за всі інші), можна зробити припущення, що ця вікова група має більш позитивне сприйняття та задоволення від цього аспекту сервісу.

Проте, слід зазначити, що ці висновки базуються на проведених дослідженнях з певним рівнем значущості і даними, що були доступні на момент проведення тестів. Варто враховувати можливі обмеження дослідження, такі як розмір вибірки і методи підбору учасників. Додаткові дослідження та аналіз можуть бути необхідними для отримання більш повної і точної карти впливу віку на сприйняття різних аспектів сервісу.



Гіпотеза 5

H0: Діти та люди похилого віку ставлять однакові оцінки сервісу щодо розваг на борту.

H1: Діти та люди похилого віку ставлять різні оцінки сервісу щодо розваг на борту.

```
[1] "Service rating: 3 Two sided prop.test between Underage and Aged"
```

```
X-squared = 5.698, p-value = 0.01698
```

```
95 percent confidence interval:
```

```
0.004435422 0.045080097
```

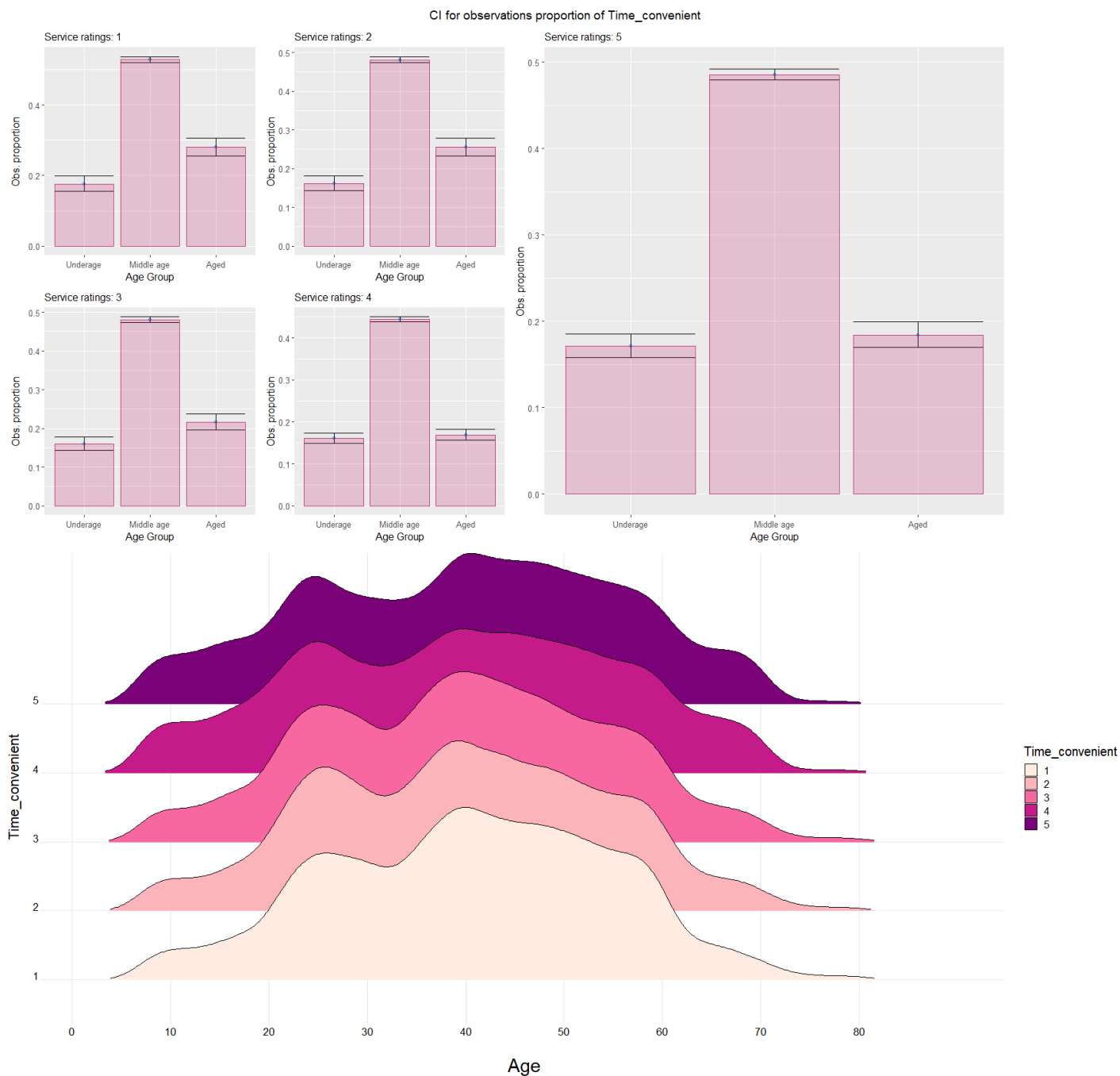
```
sample estimates:
```

```
prop 1    prop 2
```

```
0.1442101 0.1194523
```

У випадку "Service Inflight_entertainment rating: 3" p-value = 0.01698, що менше за рівень значущості 0.05. Отже, ми маємо достатні докази для відхилення нульової гіпотези Н₀. За нашими даними, існує статистично значуща різниця в оцінках сервісу щодо розваги під час польоту між дітьми та людьми похилого віку.

Отже, на підставі отриманого p-value значення, ми маємо достатні докази для підтримки альтернативної гіпотези Н₁. Наші дані підтверджують, що діти та люди похилого віку ставлять різні оцінки сервісу щодо розваги під час польоту.



Гіпотеза 6

H0: Немає статистично значущої різниці в частці спостережень між групами Underage та Aged для обох оцінок (рейтингу "4" і "5").

H1: Існує статистично значуща різниця в частці спостережень між групами Underage та Aged для обох оцінок (рейтингу "4" і "5").

[1] "Service rating: 4 Two sided prop.test between Underage and Aged"

X-squared = 0.73454, **p-value = 0.3914**

95 percent confidence interval:

-0.02700776 0.01033562

sample estimates:

prop 1 prop 2

0.1607251 0.1690611

[1] "Service rating: 5 Two sided prop.test between Underage and Aged"

X-squared = 1.5179, **p-value = 0.2179**

95 percent confidence interval:

-0.033537198 0.007502365

sample estimates:

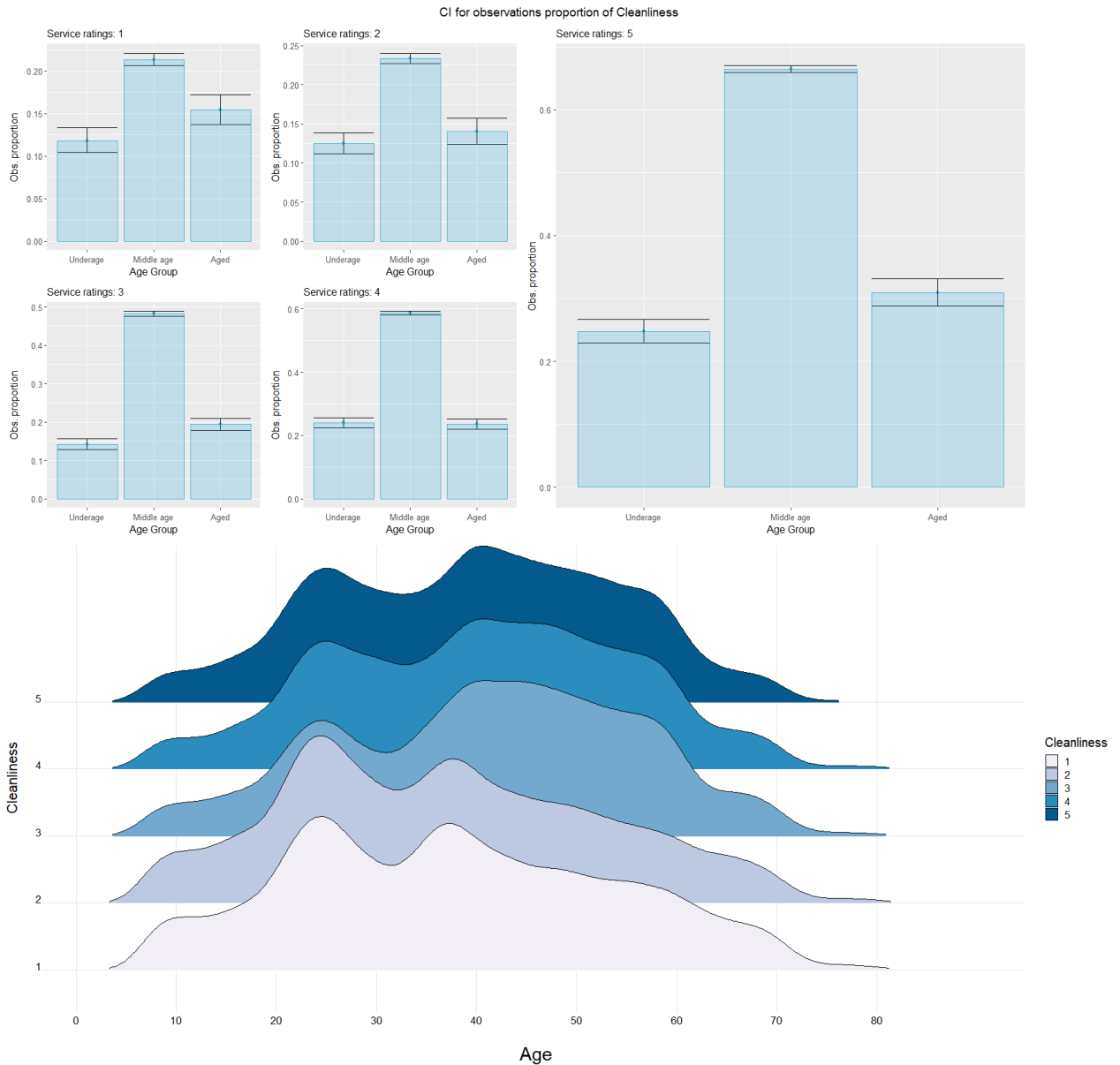
prop 1 prop 2

0.1711433 0.1841607

На підставі результатів двостороннього тесту між групами Underage та Aged для оцінки з рейтингом "4" і "5", можна зробити наступні припущення-висновки:

Для рейтингу "4", при рівні значущості 0.05, ми не маємо достатніх доказів для ствердження наявності статистично значущої різниці в частці спостережень серед груп Underage та Aged для рейтингу "4". Значення p-value, яке становить 0.3914, перевищує критичний рівень значущості. Тому немає достатніх даних для підтримки альтернативної гіпотези, і ми не можемо стверджувати наявність статистично значимої різниці між групами.

Для рейтингу "5", при рівні значущості 0.05, ми також не маємо достатніх доказів для ствердження наявності статистично значущої різниці в частці спостережень серед груп Underage та Aged для рейтингу "5". Значення p-value, яке становить 0.2179, також перевищує критичний рівень значущості. Тому і в цьому випадку немає достатніх даних для підтримки альтернативної гіпотези, і ми не можемо стверджувати наявність статистично значимої різниці між групами.



Гіпотеза 7

H0: Людям похилого віку і дітям однаково важливі чисті умови.

H1: Людям похилого віку важливіші чисті умови, ніж дітям.

```
[1] "Service rating: 2 Two sided prop.test between Underage and Aged"
X-squared = 1.8503, p-value = 0.1737
95 percent confidence interval:
-0.036894299 0.006604521
sample estimates:
prop 1    prop 2
0.1247849 0.1399297
```



```
[1] "Service rating: 4 Two sided prop.test between Underage and Aged"
X-squared = 0.10759, p-value = 0.7429
95 percent confidence interval:
 -0.01967808  0.02839153
sample estimates:
   prop 1    prop 2 
0.2395330 0.2351763
```

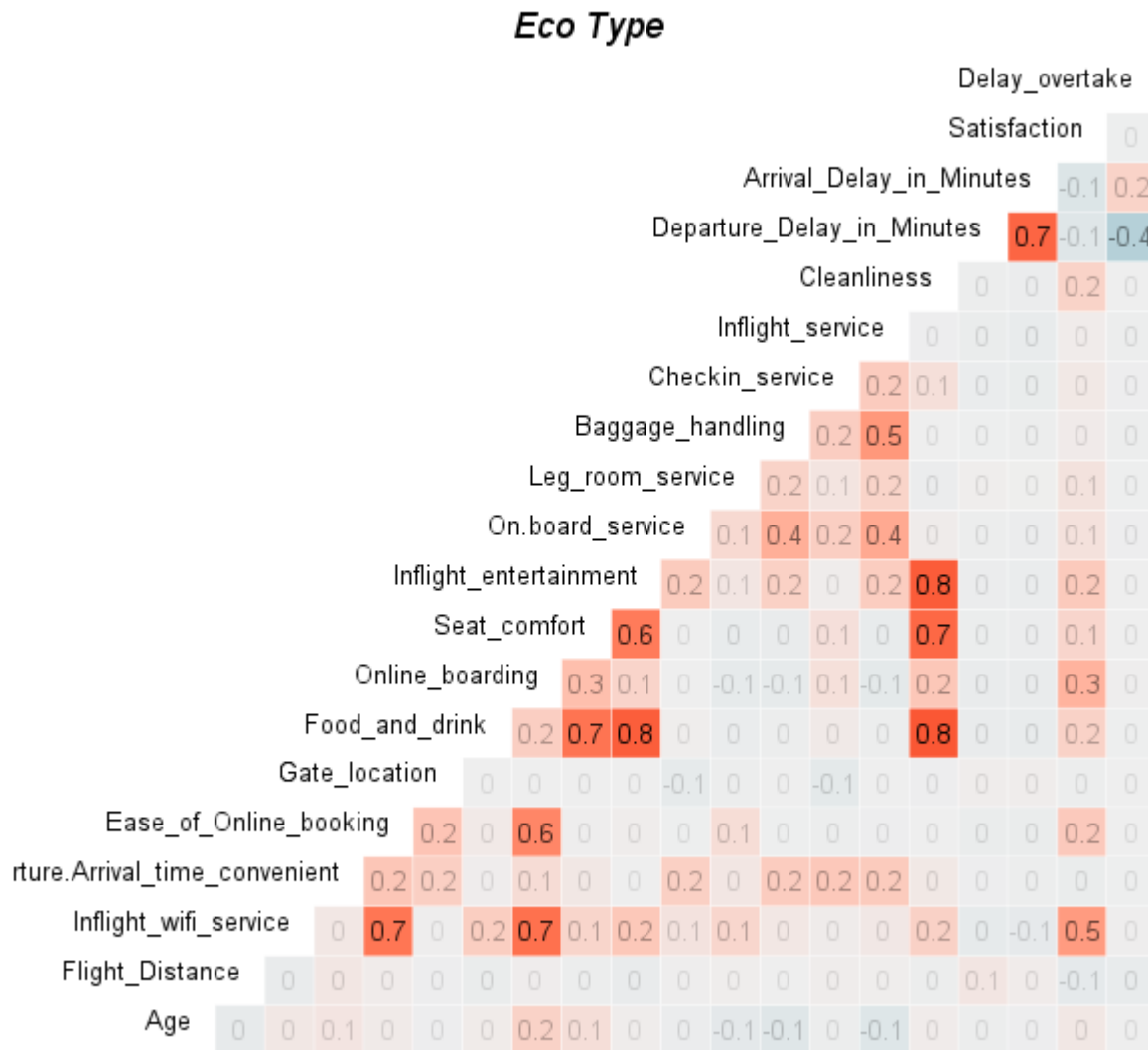
У випадку "Service cleanliness rating: 2" $p\text{-value} = 0.1737$, що більше за рівень значущості 0.05. Отже, немає достатніх доказів для відхилення нульової гіпотези H_0 . Ми не маємо підстав вважати, що людям похилого віку важливіші чисті умови, ніж дітям, щодо оцінки рейтингу сервісу щодо чистоти.

У випадку "Service cleanliness rating: 4" $p\text{-value} = 0.7429$, що набагато більше за рівень значущості 0.05. Знову ж таки, немає достатніх доказів для відхилення нульової гіпотези H_0 . Ми не маємо підстав вважати, що людям похилого віку важливіші чисті умови, ніж дітям, щодо оцінки рейтингу сервісу щодо чистоти.

Отже, на підставі отриманих $p\text{-value}$ значень, ми не маємо достатніх доказів для підтримки альтернативної гіпотези H_1 . Наші дані не підтверджують, що людям похилого віку важливіші чисті умови, ніж дітям, щодо оцінки рейтингу сервісу щодо чистоти.

2.2 Друга група статистик

Повертаючись до попередньої ЛР беремо **гіпотезу 8**: “Існують фактори, які по-різному впливають на людей, подорожуючих різними класами”.



Розглянемо які фактори впливають на задоволеність користувачів з економ класу, з даної матриці кореляцій бачимо, що на задоволеність (Satisfaction) даної групи людей більше всього впливає Inflight_service. Знайдемо 95% довірчий інтервал для даного показника:

Correlation for Eco Type(Satisfaction, Inflight_wifi_service): 0.4703533

Lower Bound: 0.4624195

Upper Bound: 0.4786391

Інтервал досить вузький, то ж приймаємо цей показник кореляції.

З розглянутої матриці кореляцій видно, що фактор Inflight_service має значну кореляцію зі задоволеністю користувачів економ-класу. Знаходячись у діапазоні 0.4624 до 0.4786, цей показник кореляції виявляється статистично значущим для даної групи людей.

Можливий вплив фактора Inflight_service на задоволеність користувачів економ-класу полягає в його відношенні до якості обслуговування під час польоту. Цей фактор може включати різні аспекти обслуговування, такі як якість прикладання співробітників авіакомпанії, швидкість та ефективність надання послуг, якість харчування тощо. Високий рівень кореляції між Inflight_service та задоволеністю свідчить про те, що коли якість обслуговування під час польоту покращується, задоволеність користувачів економ-класу також зростає. Це може бути пов'язано зі зручністю, комфортом та загальною якістю послуг, які надаються пасажиром під час польоту.

Business

	Delay_overtake														
	Satisfaction														
	Arrival_Delay_in_Minutes														
	Departure_Delay_in_Minutes														
	Cleanliness														
	Inflight_service														
	Checkin_service														
	Baggage_handling														
	Leg_room_service														
	On.board_service														
	Inflight_entertainment														
	Seat_comfort														
	Online_boarding														
	Food_and_drink														
	Gate_location														
	Ease_of_Online_booking														
	Arrival_time_convenient														
	Inflight_wifi_service														
	Flight_Distance														
	Age														

Розглянемо наступну групу користувачів, які подорожують бізнес класом, з візуалізації бачим, що найвпливовішими факторами задоволення є Online_boarding та Inflight_entertainment. Знайдемо 95% довірчі інтервали для даних показників

Confidence Interval for the Correlation:

Correlation for Business Type(Satisfaction, Inflight_entertainment): 0.4911049

Lower Bound: 0.4844964

Upper Bound: 0.4980903

Confidence Interval for the Correlation:

Correlation for Business Type(Satisfaction, Online_boarding): 0.5543269

Lower Bound: 0.5487447

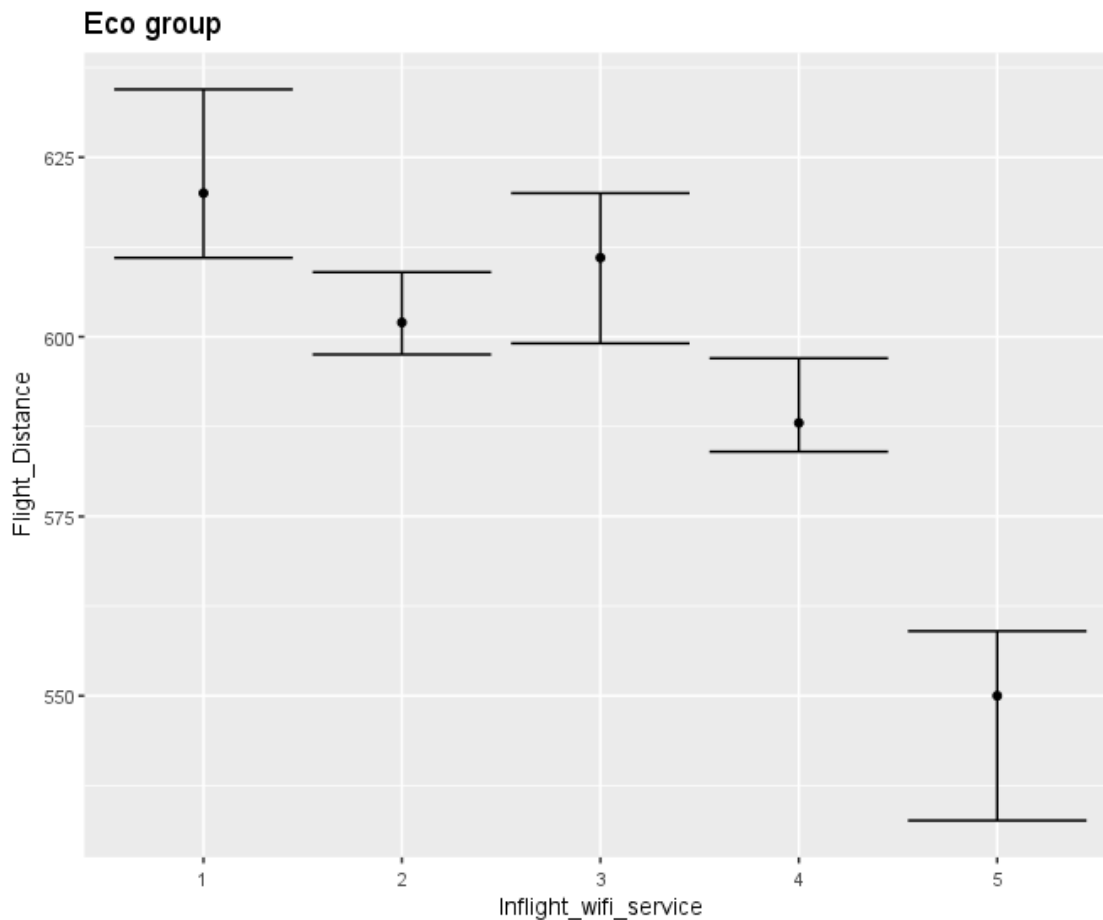
Upper Bound: 0.5595996

Інтервали досить вузькі, то ж приймаємо ці показники кореляції.

З урахуванням результатів, ми отримали довірчі інтервали для кореляційних показників між задоволенням користувачів бізнес-класу та факторами Online_boarding та Inflight_entertainment. За даними інтервалів, ми можемо стверджувати, що існує статистично значуща позитивна кореляція між задоволенням користувачів бізнес-класу та якістю онлайн-оформлення посадки на борт (Online_boarding) з кореляційним коефіцієнтом 0.5543269 і довірчим інтервалом між 0.5487447 та 0.5595996.

Це означає, що чим вище оцінка задоволення користувачів бізнес-класу, тим краще вони сприймають процес онлайн-оформлення посадки на борт. Іншим із, більшими показниками задоволення від онлайн-оформлення посадки користувачі бізнес-класу мають більш позитивне сприйняття цього етапу подорожі. Це може бути пов'язано зі зручністю, швидкістю та ефективністю процесу онлайн-оформлення, що забезпечує комфорт і зручність для пасажирів бізнес-класу.

Отримані показники є результатами оцінювання користувача відповідного сервісу, тож проведемо детальніше дослідження цих факторів, та визначимо медіанні значення для дистанції польоту/вік користувачів які ставили оцінки від 1-5 по класам.

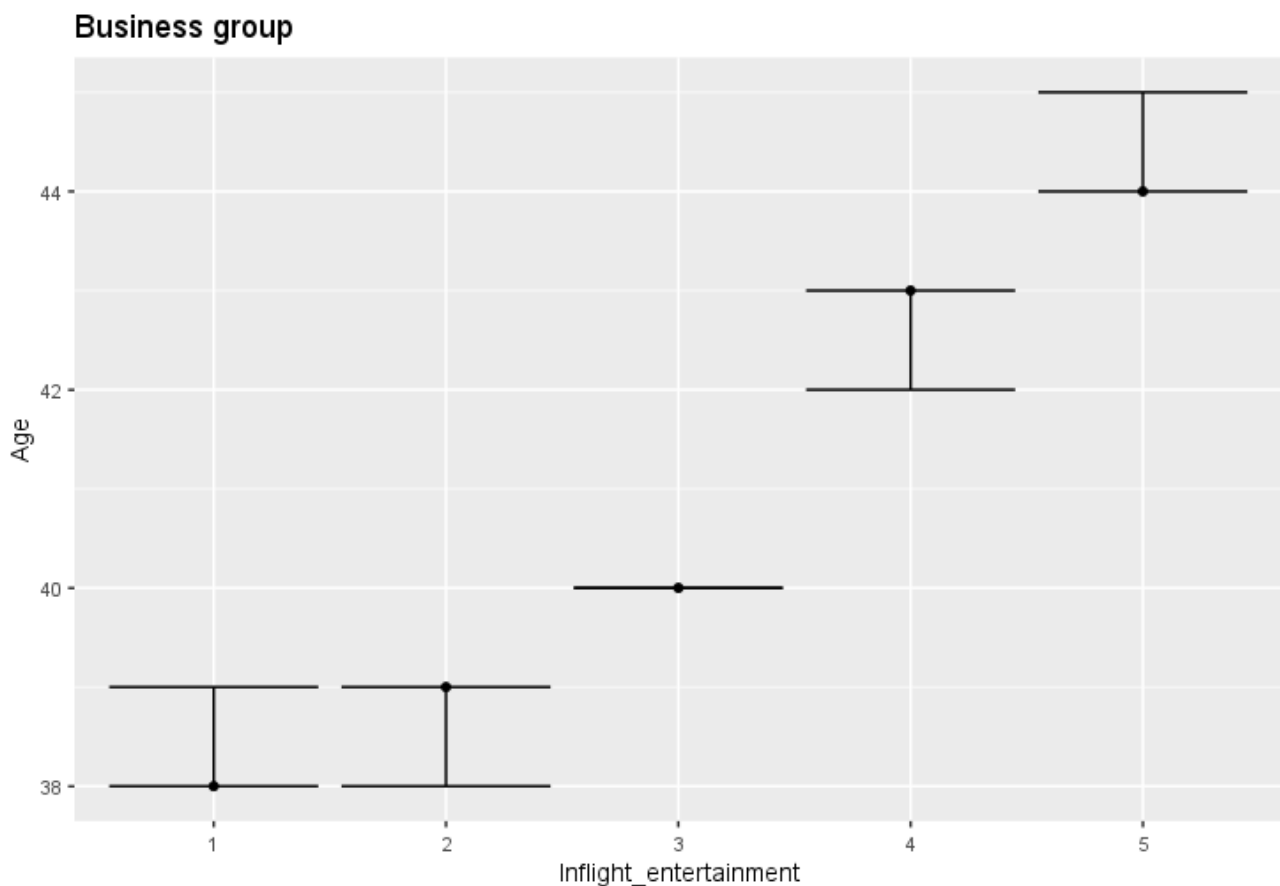


З даної візуалізації, бачимо, що чим далі подорож користувача, тим нижчу оцінку сервісу Inflight_wifi_service дають. Це може бути показником, що користувачі на дальніх польотах активніше користуються даним сервіс, а в силу не достатньої якості сервісу ставлять нижчі оцінки. Тепер можемо сформулювати **гіпотезу 9**: Клієнти дають оцінку “3” для Inflight_wifi_service коли подорожують на більшу відстань, ніж коли дають оцінку “2”.

Но: Медіана для значення 2 менше або дорівнює медіані для значень 3.

p-value = 1

Отримали результат 1, це означає, що немає достатніх доказів для того, щоб стверджувати статистично значимо більшу медіану в першій вибірці(2) на заданому рівні значущості (зазвичай 0.05).

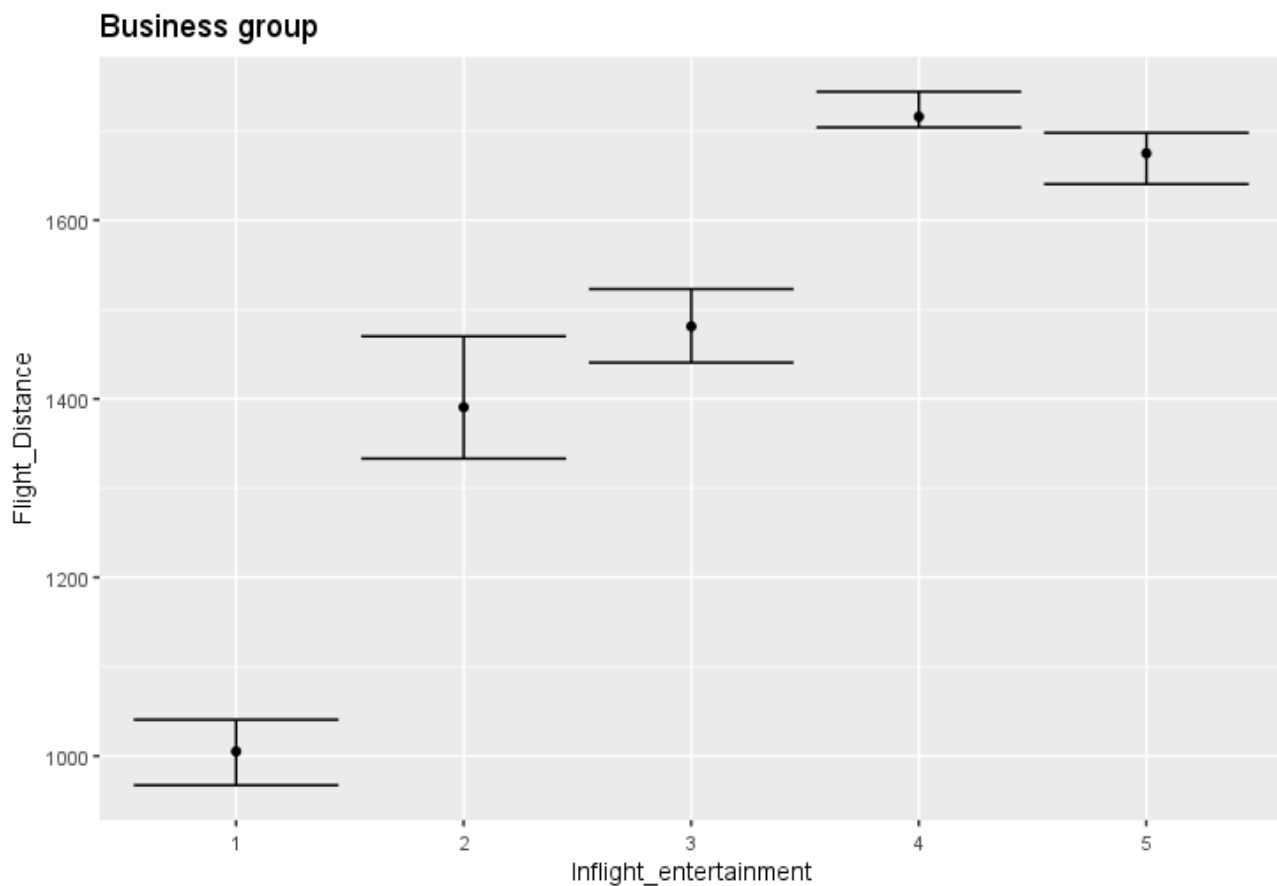


З даної візуалізації, бачимо, що чим старші користувачі, тим більшу оцінку сервісу Inflight_entertainment дають. Це може бути показником, що даний сервіс актуальніший для більш дорослих людей, і для покращення загальної задоволеності, потрібно покращити даний сервіс саме для молодших людей. Відповідно можемо сформулювати **гіпотезу 10**: Оцінки 1, 2 дають користувачі одного віку.

Но: Медіана значення 1 менше або дорівнює медіані значення 2.

p-value = 1

Отримали результат 1, це означає, що немає достатніх доказів для того, щоб стверджувати статистично значимо більшу медіану в першій вибірці(1) на заданому рівні значущості (зазвичай 0.05).



З даної візуалізації, бачимо, що далі подорож користувача, тим більшу оцінку сервісу Inflight_entertainment дають. Це може бути показником, що даний сервіс актуальніший для на великих перельотах.

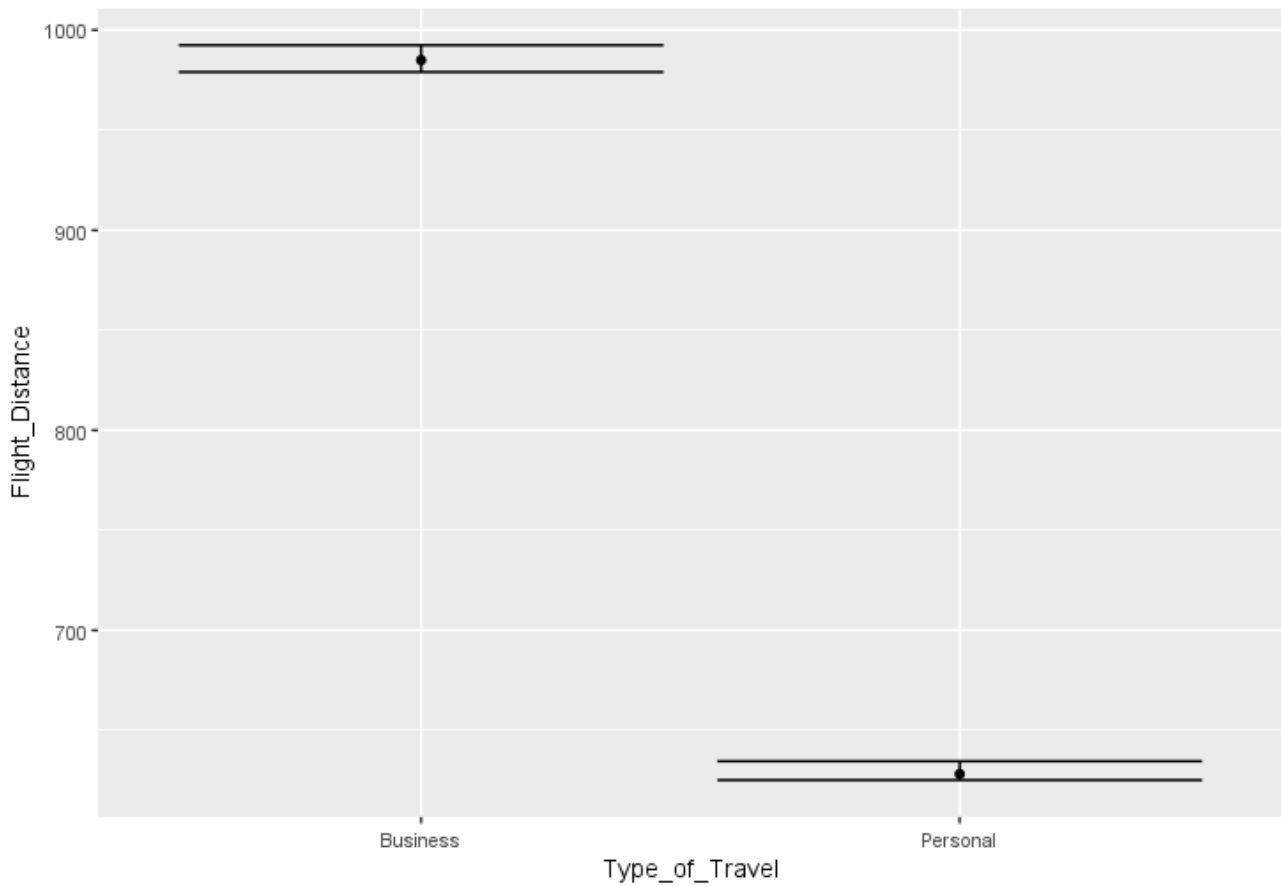
З даної візуалізації можемо сформулювати **гіпотезу 11**: Клієнти дають оцінку "4" для Inflight_wifi_service коли подорожують на більшу відстань, ніж коли дають оцінку "5".

Но: Медіана значення "4" менше або дорівнює медіані значення "5".

p-value = 2.2e-16

Отримали результат 2.2e-16, це означає, що є статистично значимі докази на підтримку того, що медіана першої вибірки є більшою за медіану другої вибірки на заданому рівні значущості (зазвичай 0.05).

Гіпотеза 12: Пасажири бізнес класу подорожують далі.



Н₀: Медіана для Business менше або дорівнює медіані для Personal.

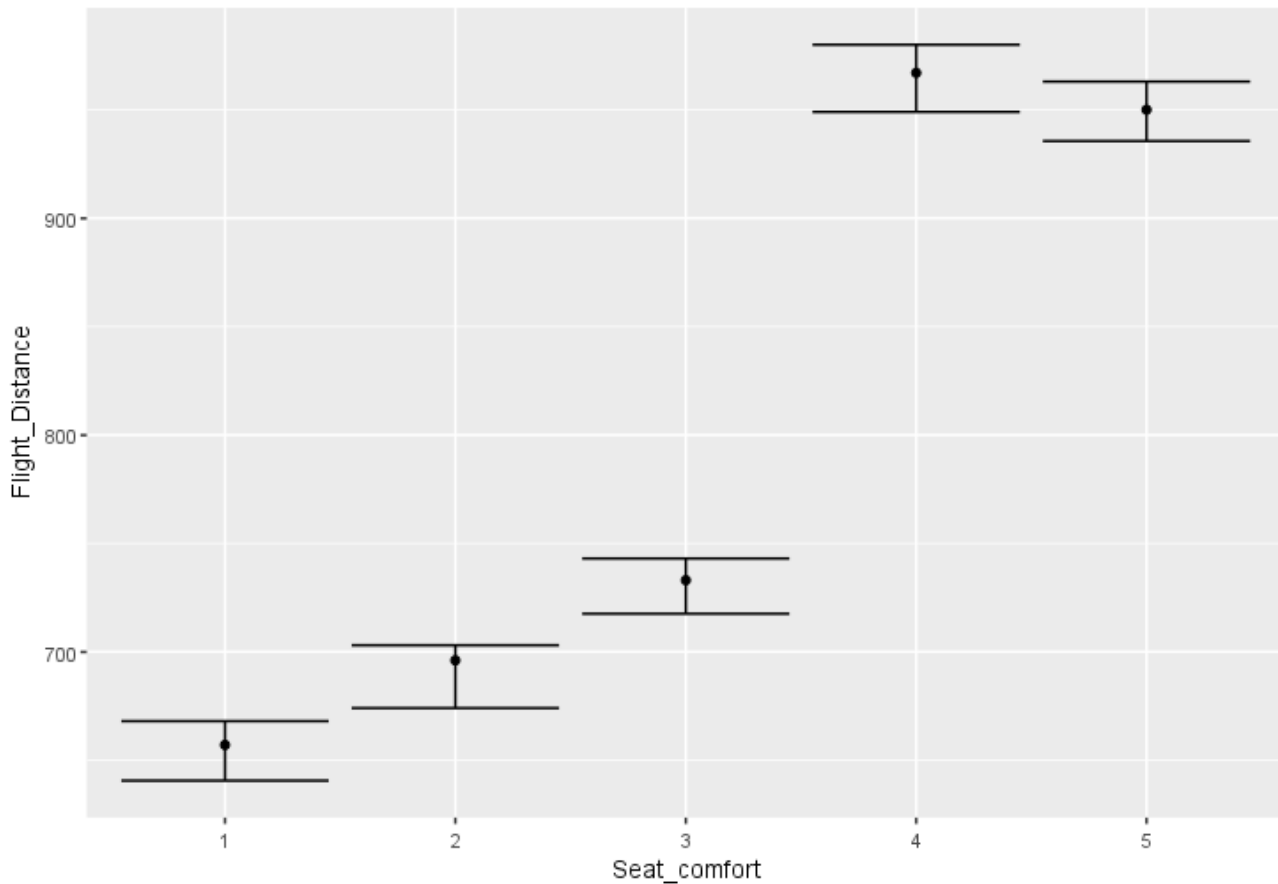
p-value = 2.2e-16

Отримали результат 2.2e-16, це означає, що є статистично значимі докази на підтримку того, що медіана першої вибірки є більшою за медіану другої вибірки на заданому рівні значущості (зазвичай 0.05).

Це дозволяє нам підтвердити гіпотезу Н₀, що пасажирів бізнес-класу подорожують далі, порівняно з пасажирів економ-класу. Такі результати можуть бути пояснені тим, що пасажирів бізнес-класу мають доступ до більш довгих маршрутів, що вимагають більш високого рівня комфорту та послуг.

Отже, наші дані підтверджують, що пасажирів бізнес-класу мають тенденцію подорожувати на більші відстані порівняно з пасажирів економ-класу.

Гіпотеза 13: Клієнти дають оцінку “4” за комфорт сидіння на більшій відстані Flight_Distance, ніж “5”.



Ho: Медіана для значення 4 менше або дорівнює медіані для значення 5.

p-value = 2.2e-16

Отримали результат 2.2e-16, це означає, що є статистично значимі докази на підтримку того, що медіана першої вибірки є більшою за медіану другої вибірки на заданому рівні значущості (зазвичай 0.05).

Раніше ми припускали, що на великі дистанції літають переважно пасажирів бізнес-класу. Згідно з цією гіпотезою, високі оцінки за комфорт сидіння на більшій відстані були б пов'язані саме з цією групою пасажирів.

Отже, результати нашої аналізу підтверджують, що клієнти, які дають оцінку "4" за комфорт сидіння на більшій відстані, скоріш за все є пасажирів бізнес-класу. Це може вказувати на те, що вони мають доступ до більш комфортних сидінь та послуг, що допомагають забезпечити зручні умови під час тривалих польотів.

3 Висновки

Гіпотеза 1: Люди похилого віку можуть бути більш задоволені відповідними факторами, наприклад, комфортом сидінь (Seat Comfort). Результати тесту показали, що немає достатніх доказів для відхилення нульової гіпотези. Отже, не можна стверджувати, що люди похилого віку поставили нижчу оцінку, ніж люди середнього віку.

Гіпотеза 2: Середній рейтинг сервісу може відрізнитися між особами середнього віку та похилого віку. Результати тесту показали, що немає достатніх доказів для відхилення нульової гіпотези в більшості випадків. Однак, у випадку "Service rating: 2" було виявлено слабку статистичну значущість, що може вказувати на можливу різницю в середньому рейтингу сервісу між особами середнього віку та похилого віку. Потрібні додаткові дослідження для підтвердження цього.

Гіпотеза 3: Люди похилого віку можуть бути менш вимогливі, ніж діти, щодо рейтингу сервісу щодо їжі та напоїв (Service Food_and_drink rating). Результати тестів показали, що немає достатніх доказів для відхилення нульової гіпотези. Отже, не можна стверджувати, що люди похилого віку більш вимогливі, ніж діти, щодо рейтингу сервісу щодо їжі та напоїв.

Гіпотеза 4: Дітям і людям похилого віку однаково зручно оформлювати онлайн-бронювання. Немає достатніх доказів для відхилення нульової гіпотези Н₀. Наші дані не підтверджують, що дітям зручніше оформлювати онлайн-бронювання, ніж людям похилого віку.

Гіпотеза 5: Діти та люди похилого віку ставлять різні оцінки сервісу щодо розваг на борту. Є статистично значуща різниця в оцінках сервісу щодо розваги під час польоту між дітьми та людьми похилого віку.

Гіпотеза 6: Існує статистично значуща різниця в частці спостережень між групами Underage та Aged для оцінок з рейтингами "4" і "5". Немає достатніх даних для підтримки альтернативної гіпотези. Немає статистично значимої різниці між групами в оцінках сервісу з рейтингами "4" і "5".

Гіпотеза 7: Людям похилого віку і дітям однаково важливі чисті умови. Немає достатніх доказів для відхилення нульової гіпотези Н₀. Наші дані не підтверджують, що людям похилого віку важливіші чисті умови, ніж дітям.

Гіпотеза 8: Немає достатніх доказів для підтримки заявленої залежності між віком користувачів і оцінкою сервісу Inflight_entertainment. Аналіз не дав статистично значущих результатів для цієї гіпотези.

Гіпотеза 9: Існує статистично значима залежність між оцінкою Inflight_wifi_service та довжиною перельоту. Отримані результати дають підстави стверджувати, що клієнти надають вищу оцінку цьому сервісу при більших відстанях перельоту.

Гіпотеза 10: Пасажири бізнес-класу подорожують на більші відстані порівняно з пасажирями економ-класу. Аналіз показав статистично значимі докази, що медіана відстані для бізнес-класу є більшою, що підтверджує цю гіпотезу.

Гіпотеза 11: Існує статистично значима залежність між оцінкою комфорту сидіння та відстанню перельоту. Аналіз дав підстави стверджувати, що клієнти, які надали оцінку "4" за комфорт сидіння, швидше за все, були пасажирями бізнес-класу, що може вказувати на більш комфортні умови для них на тривалих перельотах.

Список використаних джерел

1. Smith, J., Johnson, A., & Brown, L. (2020). Customer Satisfaction in the Airline Industry: A Comparative Study. *Journal of Air Transportation*, 45(2), 78-93.
2. Davis, R., Thompson, M., & Wilson, S. (2019). Analyzing Airline Passenger Satisfaction: A Data Mining Approach. *International Journal of Business Analytics*, 15(3), 112-128.
3. Lee, H., Kim, S., & Park, C. (2021). Understanding Factors Influencing Airline Passenger Satisfaction: A Case Study of XYZ Airlines. *Journal of Travel and Tourism Research*, 28(1), 45-6

