

INPATIENT ANALYSIS AND PREDICATING LENGTH OF STAY

Presentation for laboratory work №1

Developed by:

- Babych Iryna
- Ivanyk Yurii
- Romanetskiy Nick
- Sukhovii Ihor
- Sholop Lyubomyr

Основні дослідницькі питання

Дослідницьке питання

- I. Люди якого віку / статі / раси більше часу проводить в лікарні
- II. Який вік / стать / раса людей найчастіше зустрічається в лікарні
- III. Ризик смертності в залежності від віку / статі / раси

Попередня відповідь

- I. Тривалість перебування в лікарні буде в середньому меншою у дітей та повільно зростатиме до 30 років
- II. Серед пацієнтів буде більше людей старшого віку. Розподіли за статтю та расою мають приблизно відповідати таким самим посеред населення штату Нью-Йорк станом на 2015 рік.
- III. Ризик смертності, на нашу думку, зростатиме з віком і, скоріш за все, не залежатиме від статі.

Опис даних

- Обраний датасет містить дані щодо надання лікарських послуг пацієнтам у штаті Нью-Йорк за 2015 рік.
- Він не містить даних, які би могли вказувати на причетність до них окремих осіб. Також, з цього файлу було виключено вторинні діагнози та процедури, а також коди оплати послуг. Один рядок відповідає інформації про рівно 1 клінічний випадок.
- Вік пацієнтів представлений у таких вікових групах:
 - від 0 до 17 років
 - від 18 до 29 років
 - 30 до 49 років
 - від 50 до 69 років
 - і від 70 років і старше

Датасет перед очисткою містив:

- 2346760 записів та 37 змінних

Після очистки:

- 2342182 записів та 33 змінних

Робота з датасетом

Безпосередньо перед розвідковим аналізом даних, було виконано:

- ✓ Перевірку даних на наявність пропущених значень.
- ✓ Очищення
- ✓ Перевірку колонок на відповідність типам даних
- ✓ Обраховано дескриптивні статистики

Змінні датасету

- Кожна колонка датасету відповідає одній змінній.
- Більшість змінних в датасеті є категорійними
- Числовими і впорядкованими змінними в датасеті є:
 - *Length.of.Stay*
 - *Birth.Weight* (через природу даних містить найбільшу кількість NA значень)
 - *Total.Charges*
 - *Total.Costs*

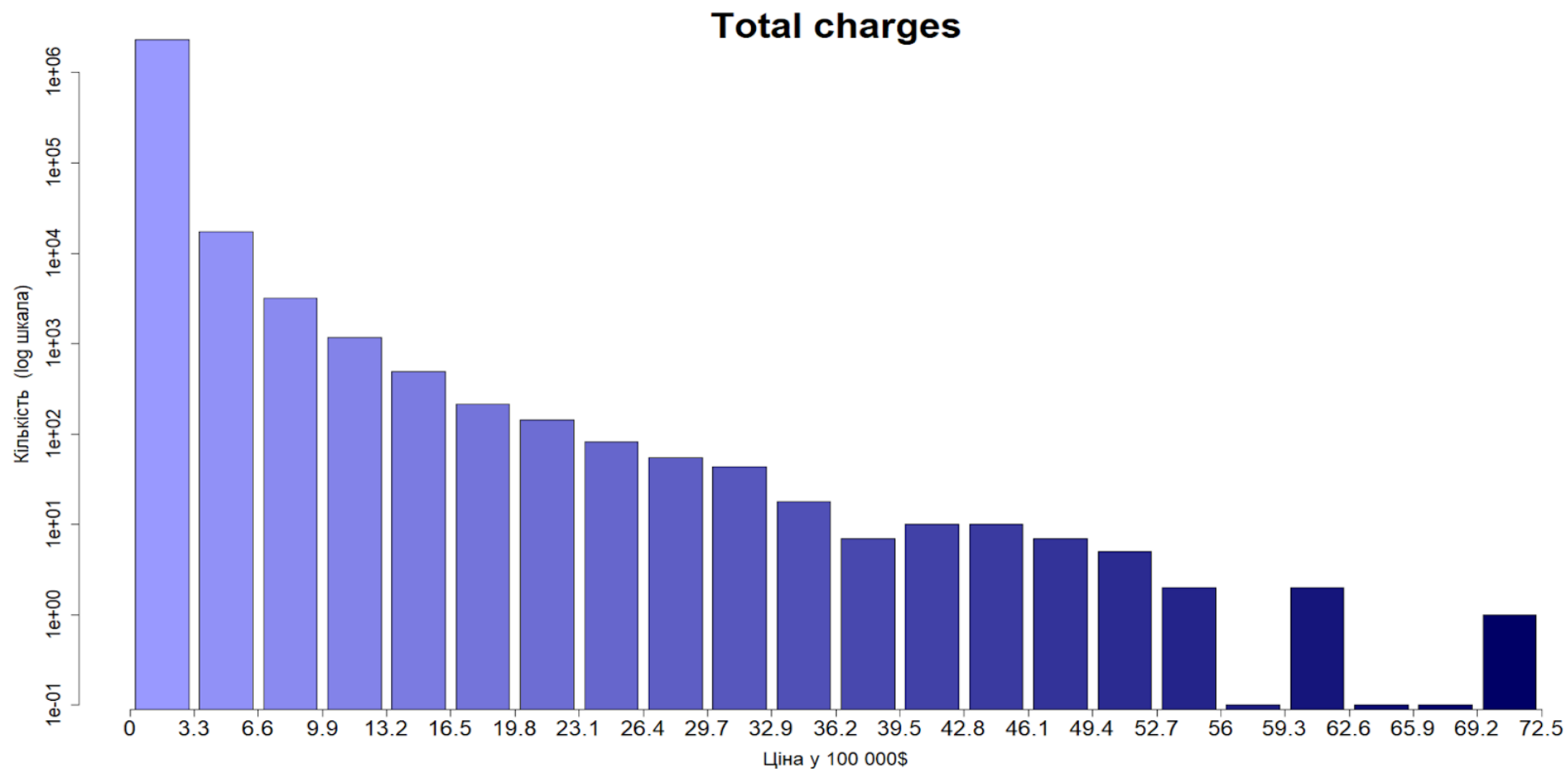
Графік тривалості перебування за log шкалою кількості випадків



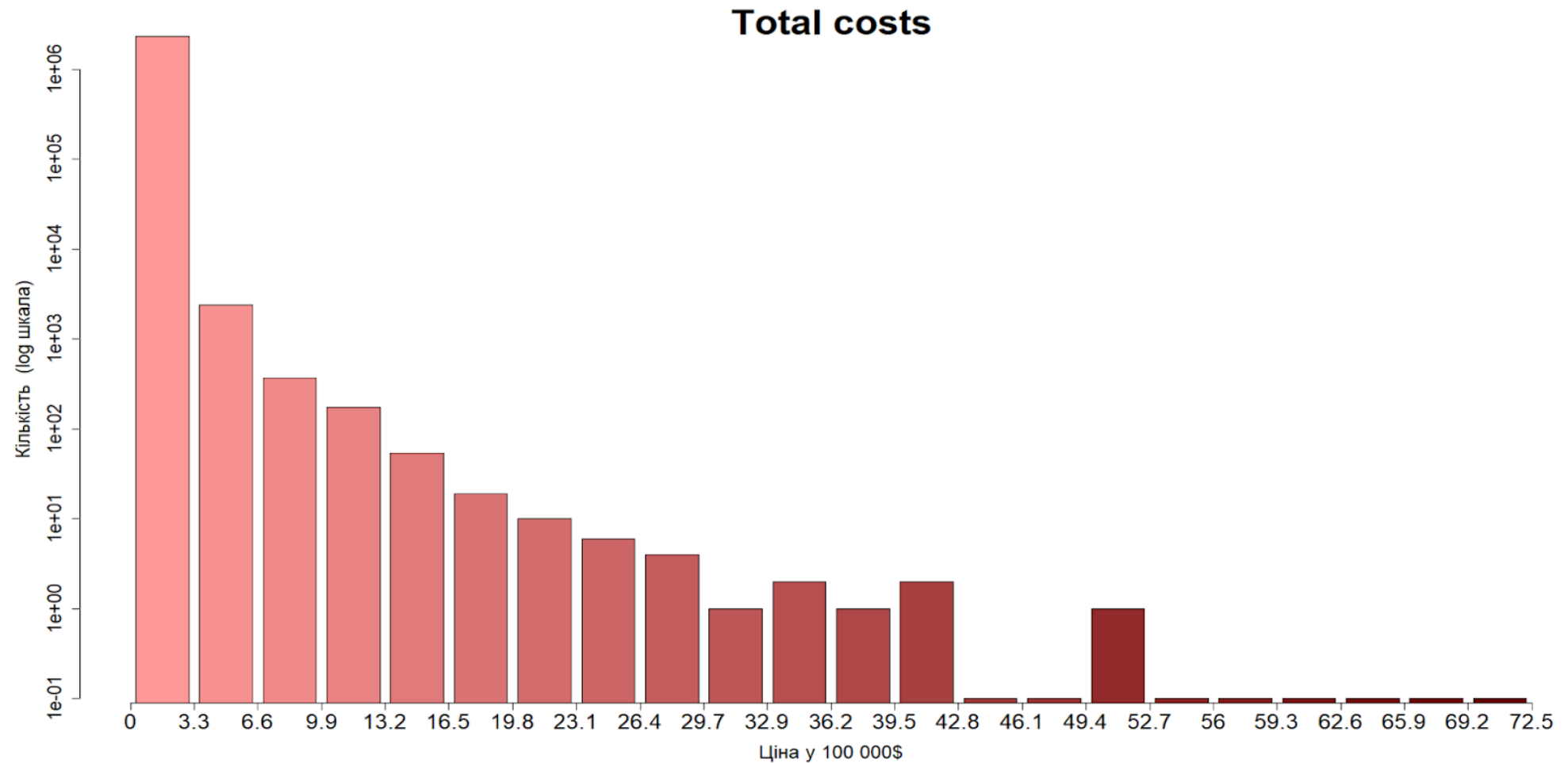
Графік логарифму привалості перебування в лікарні



Графік загальної ціни з log шкалою кількості випадків

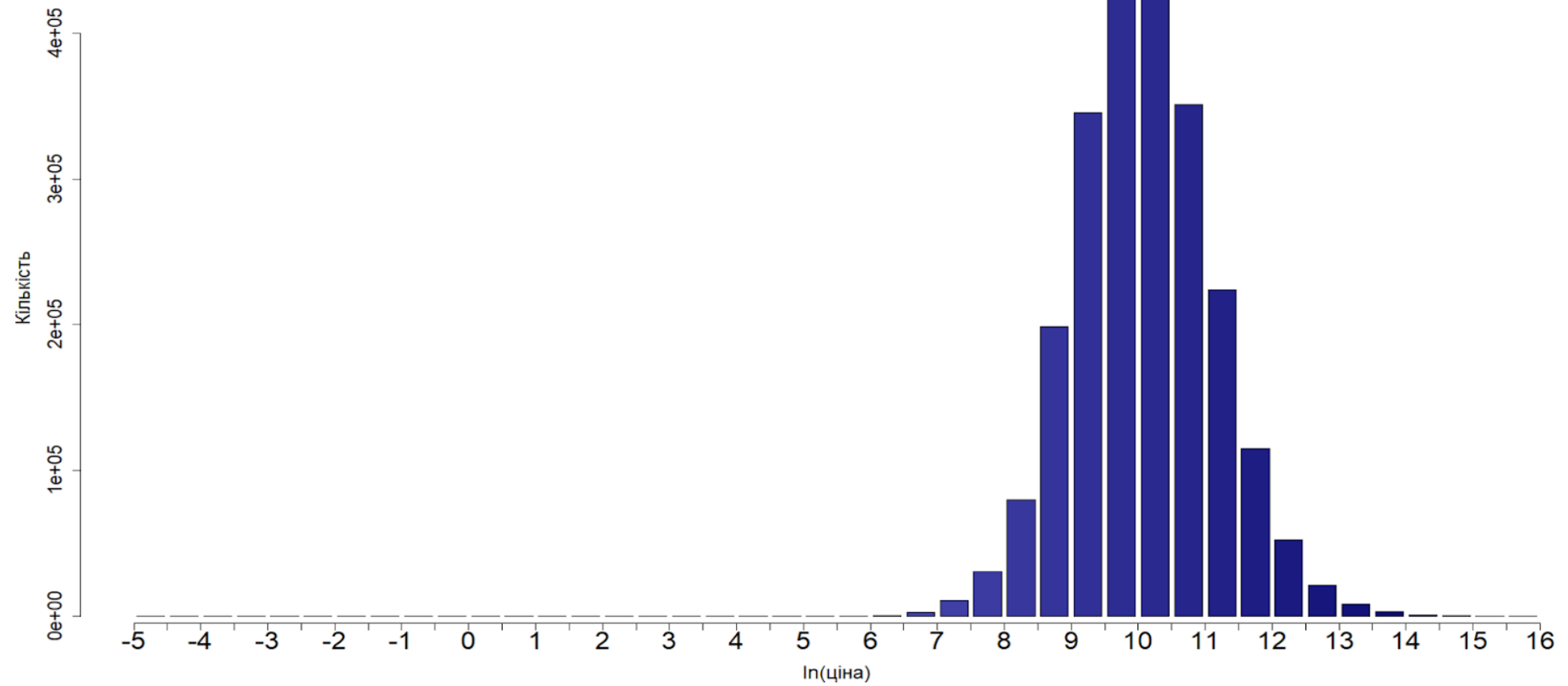


Графік загальної собівартості з log шкалою кількості випадків

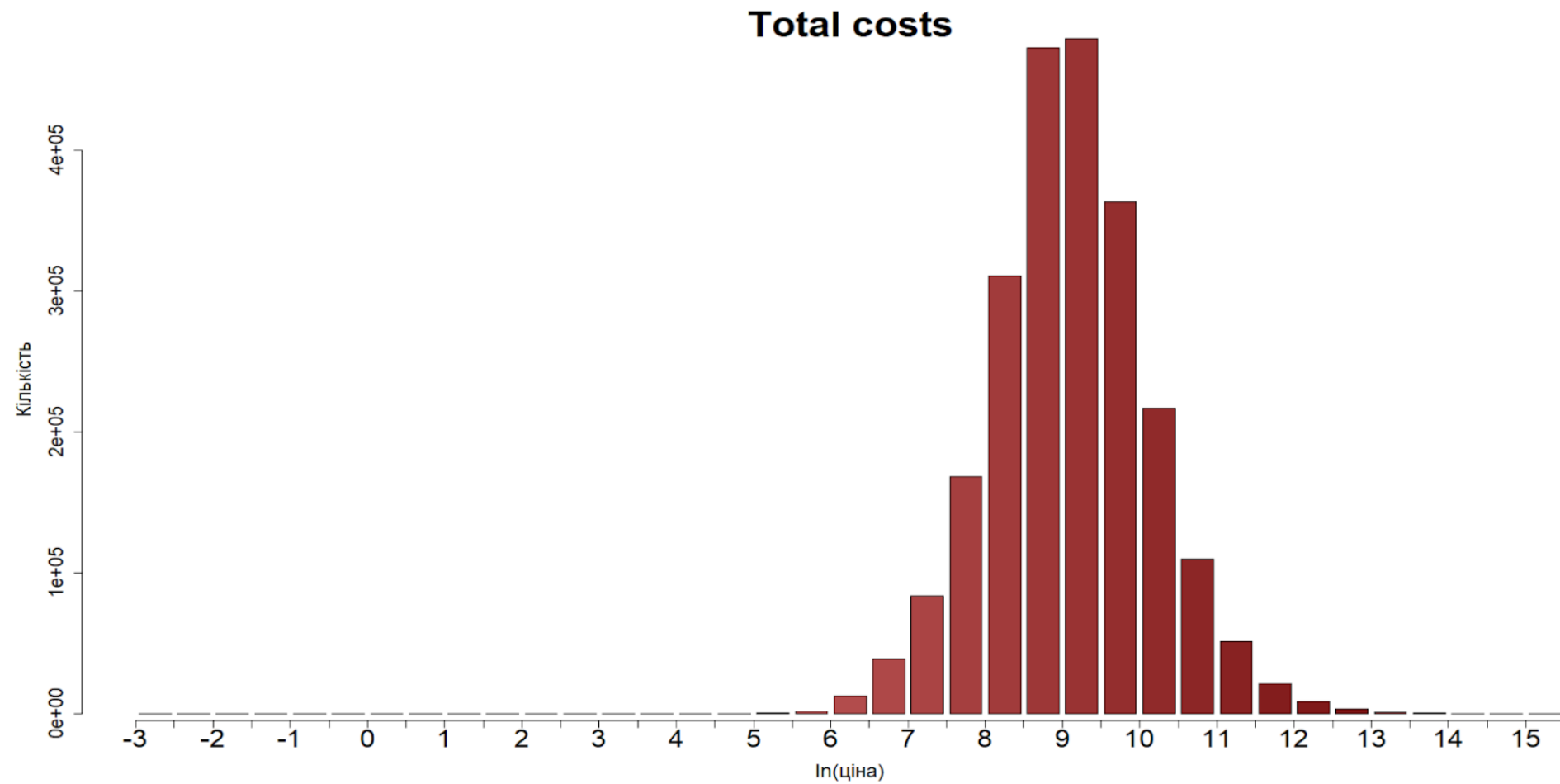


Графік логарифму ціни

Total charges



Графік логарифму собівартості

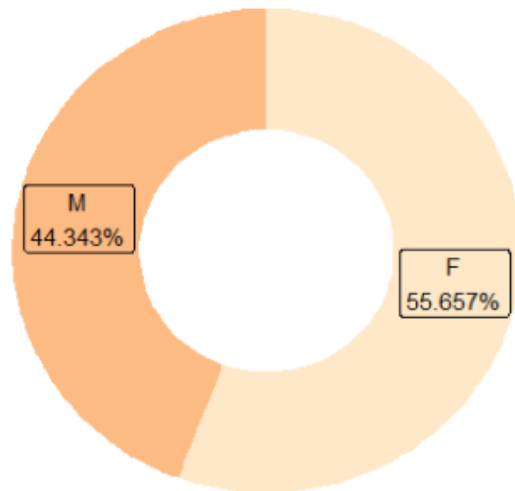


Результати EDA

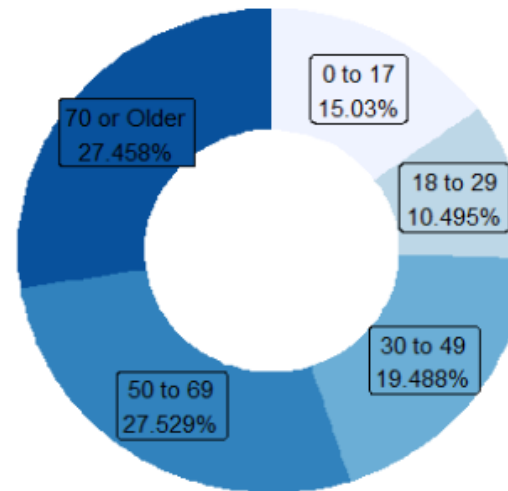
Розподіл пацієнтів за віком, расою та статтю

Частота лежання в лікарні

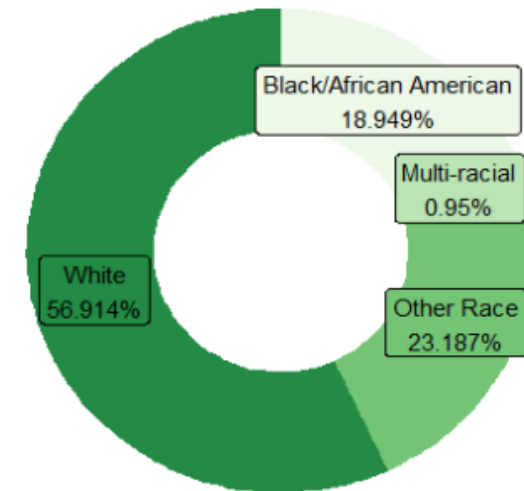
по статі



по віку



по расі

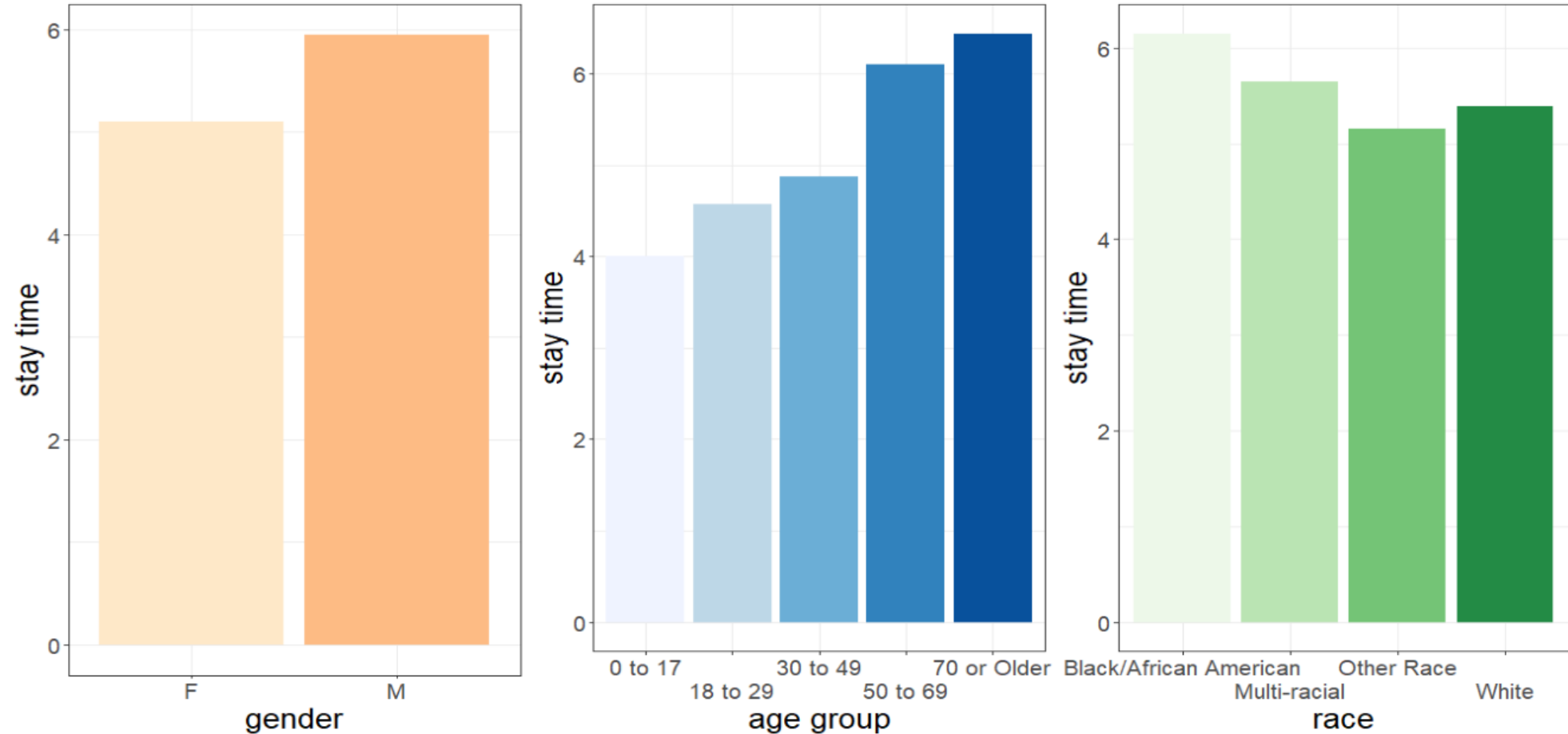


Проміжні висновки

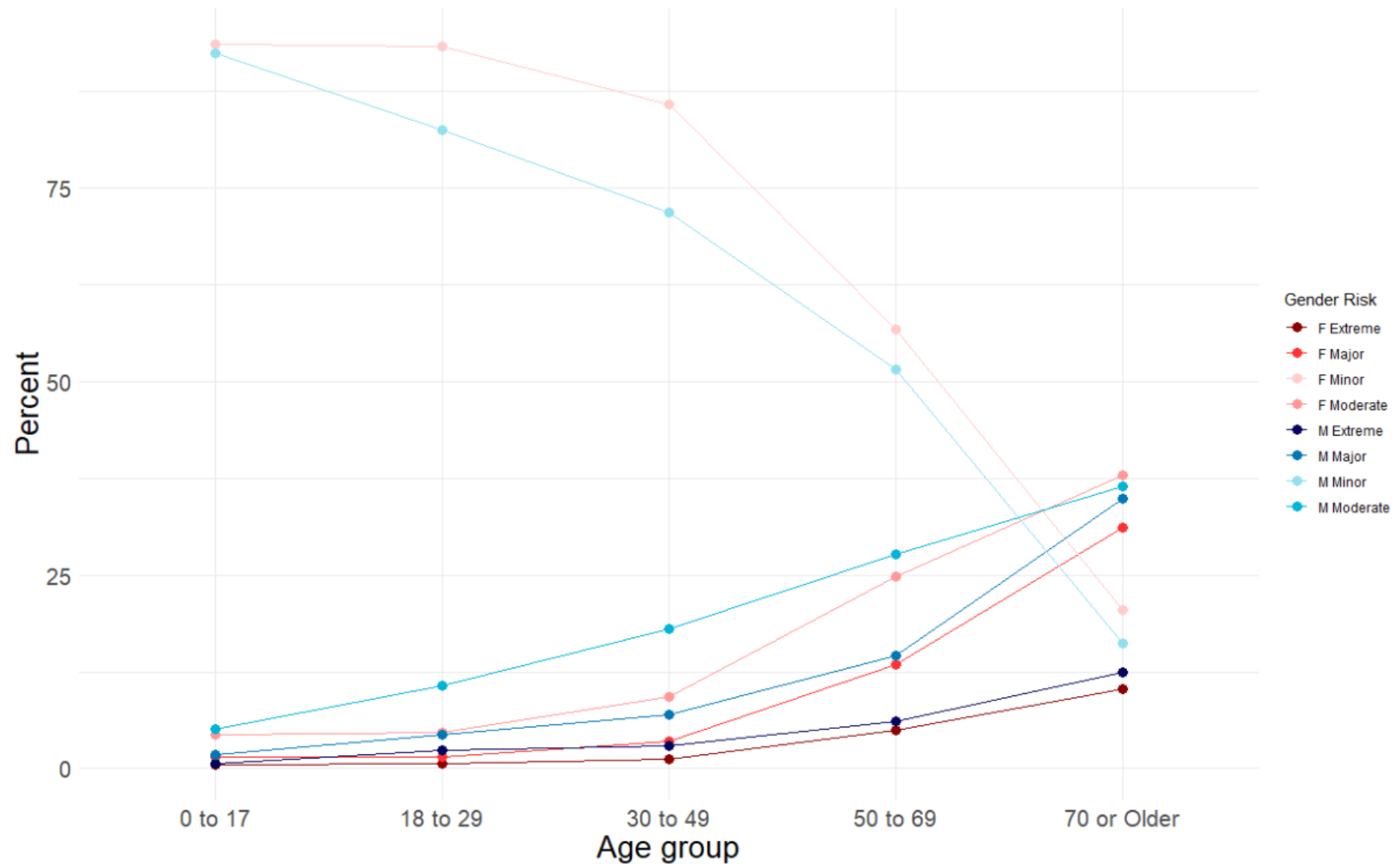
- Розбиття пацієнтів за статтю трохи відрізняється від відповідного розбиття серед населення, але різниця (приблизно 4%) невелика, отже, вона може виявитись статистично незначущою за подальшого дослідження.
- Люди, старші за 50 років, потрапляють в лікарню частіше за інших, оскільки їх відсоток серед пацієнтів переважає їх відсоток серед населення. Особливо серед них вирізняються пацієнти 70+, оскільки в них відсоток різниці між кількістю населення і кількістю пацієнтів найбільший. Серед молоді і людей середніх років різниця даними з датасету та даними про населення приблизно однакова.
- Найменше відносно їх частки в популяції в лікарню потрапляють білі пацієнти, а найбільше - представники інших рас, тобто азіати, корінні американці та жителі Тихоокеанських островів. Пояснення, чому так відбувається, потребує подальшого дослідження.

Залежності тривалості лікування від віку, статі та раси

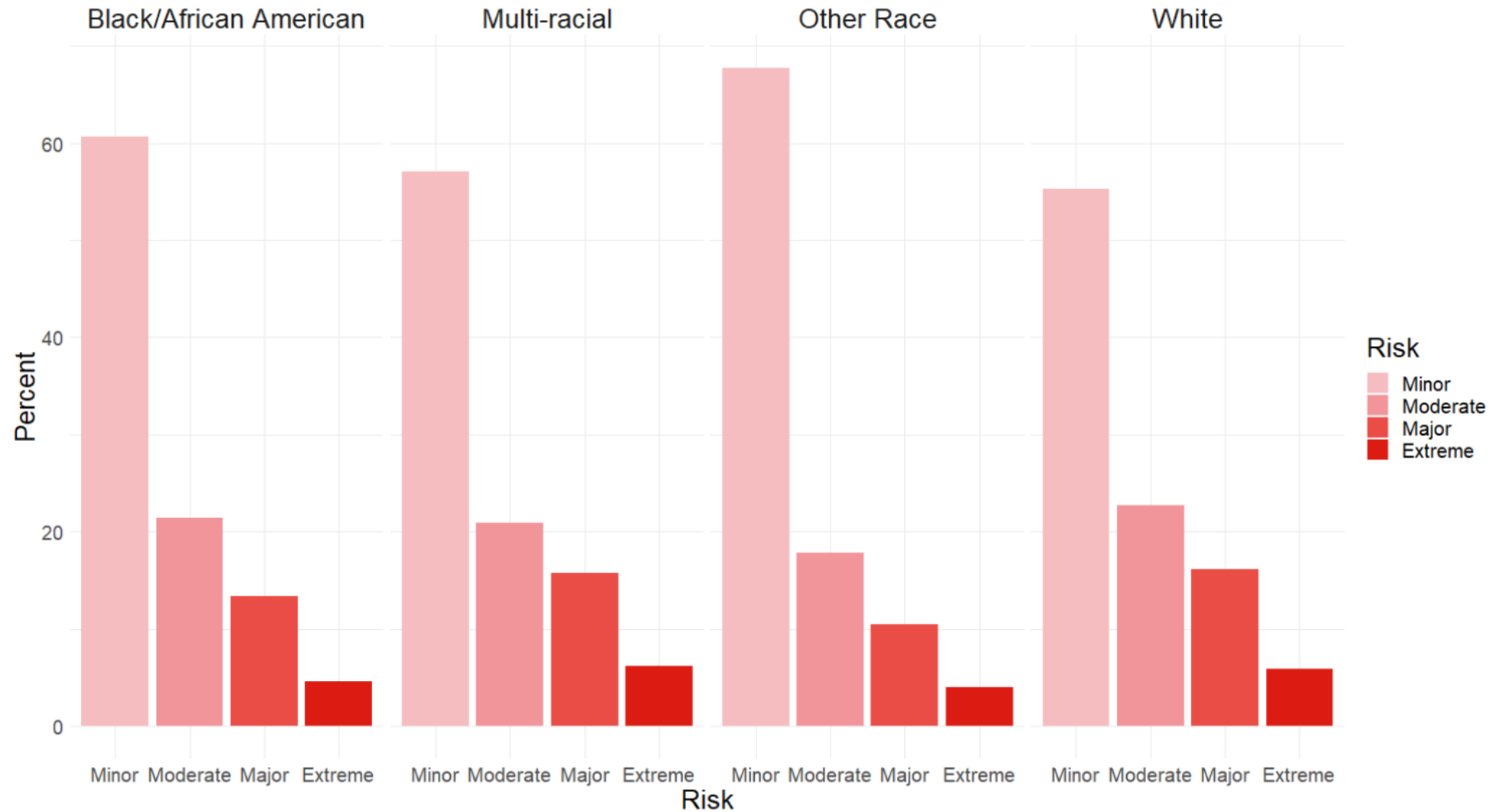
Середня тривалість перебування в лікарні



Ризик смертності у відсотках в залежності від віку та статі



Ризик смертності в залежності від раси



Висновки

1. Розподіл собівартостей та цін за обслуговування в лікарнях схожий на логарифмічно нормальний. Розподіл привалостей перебування в лікарні не схожий на якісь специфічні, а кількість випадків спадає за збільшення привалості
2. Щодо характеристик пацієнтів за віком: їх розподіл відносно кількості людей цієї ж категорії в населенні дає підґрунтя для спостереження, що люди до 50 років потрапляють в лікарню приблизно з однаковою частотою, для 50-69 років ця частота зростає, а найчастіше туди потрапляють люди віку 70+ років, якщо дані різниці є статистично значущими. **Це відповідає попереднім припущенням про те, що в лікарню здебільшого потрапляють люди старшого віку.**
3. **Чоловіки менше потрапляють в лікарню**, при чому, навіть якщо врахувати їх відсоток в населенні. **Наші припущення стверджували, що характеристики за статтю будуть приблизно однакові і в чоловіків, і в жінок.** Це припущення, здається, доведеться відкинути, але лише після подальшої перевірки.
4. Представники **білої раси потрапляють в лікарні менше**, ніж можна було б очікувати за розподілом населення, а представники **інших рас (люди азійського походження та корінні жителі Америки)- більше**, що відрізняється від нашої гіпотези