

НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ УНІВЕРСИТЕТ УКРАЇНИ

«КИЇВСЬКИЙ ПОЛІТЕХНІЧНИЙ ІНСТИТУТ

ІМЕНІ ІГОРЯ СІКОРСЬКОГО»

Факультет прикладної математики

Кафедра прикладної математики

Лабораторна робота №2

із дисципліни “Аналіз даних”

на тему:

“Статистичне виведення”

Виконали:

команда

“CrAzy_pEnGuIn_cOdeRs”

Атаман Юлія, КМ-11

Кракович Борислав, КМ-11

Лушников Іван, КМ-11

Шушпаннікова Інна, КМ-11

Ягода Євгенія, КМ-11

Керівник:

старший викладач

Тавров Д.Ю.

Київ — 2024

Зміст

ВСТУП	3
Обчислені статистики	6
Графіки з лабораторної 1 та довірчі інтервали	6
Гіпотеза 1. Про рівність середньої кількості фатальних аварій в теплі та холодні місяці.	22
Гіпотеза 2. Про рівність середньої кількості фатальних аварій у робочі та вихідні дні.	25
Гіпотеза 3. Молодші водії більш схильні до аварій через їхню більшу схильність до ризикованої поведінки та нестачу досвіду порівняно зі старшими водіями.(маневри і т.д).	28
Гіпотеза 4. Про рівність фатальних аварій автомобілей з більшим та меншим віком за роками	30
Гіпотеза 5. Про рівність середньої кількості за роками фатальних аварій у міській та сільській місцевості .	32
Протестовані гіпотези	33
Гіпотеза 1. Про рівність середньої кількості аварій в теплі та холодні місяці.	3334
Гіпотеза 2. Про рівність середньої кількості фатальних аварій у робочі та вихідні дні.	35
Гіпотеза 3. Молодші водії більш схильні до аварій через їхню більшу схильність до ризикованої поведінки та нестачу досвіду порівняно зі старшими водіями.(маневри і т.д).	36
Гіпотеза 4. Про рівність фатальних аварій автомобілей з більшим та меншим віком за роками.	37
Гіпотеза 5. Про рівність середньої кількості за роками фатальних аварій у міській та сільській місцевості .	40
Висновки	42
Список літератури	43

ВСТУП

Мотивація дослідження

Мотивація роботи полягає у дослідженні інформації з бази даних про дорожньо-транспортні пригоди у Великобританії в період з 2006 по 2015 роки з метою їх аналізу та встановлення зв'язків між характеристиками транспортних засобів, водіїв, дорожніх покриттів тощо та власне ДТП для виявлення факторів, що можуть бути причинами ДТП. Результати роботи можна використати як інформаційні матеріали, що можуть:

- підвищити обізнаність водіїв і пішоходів, що допоможе запобігти аваріям
- бути передані і використані відповідними організаціями для покращення ситуацій на дорогах
- дати інформацію про недоліки або проблеми деяких транспортних засобів, що можуть призвести до ДТП.

Дослідницькі питання

1. Як змінюється кількість аварій залежно від сезону?
2. Чи відрізняється частота фатальних аварій у робочі та вихідні дні?
3. Чи впливає стать водія на частоту фатальних аварій?
4. Як впливає вік водія на ймовірність участі в аварії?
5. Чи пов'язаний вік автомобіля з ризиком фатальних аварій?
6. Які дорожні умови частіше призводять до фатальних аварій?
7. Чи відрізняється безпека дорожнього руху в сільській і міській місцевості?

Ми вирішили підібрати нові питання для звіту з кількох важливих причин:

Забезпечення глибини аналізу: Нові питання дозволяють зосередитися на більш глибокому та комплексному розумінні факторів, що впливають на дорожньо-транспортні події. Це включає вивчення взаємозв'язків між різними аспектами дорожнього руху, такими як умови водіння, характеристики автомобілів, особливості водіїв та зовнішні умови.

Збалансованість дослідження: Переформулювання питань дозволяє забезпечити більш збалансований підхід до аналізу причин та наслідків

аварій. Це допомагає уникнути зосередження уваги тільки на одному аспекті та забезпечує розгляд ширшого спектру чинників.

Запобігання аваріям: Глибше розуміння причин аварій через комплексний аналіз нових питань може допомогти у запобіганні майбутніх аварій. Це включає ідентифікацію ризиків, які можуть бути зменшені або виправлені через інженерні, освітні або законодавчі зміни.

Цей процес не тільки покращує якість дослідження, але й допомагає у виробленні практичних рекомендацій, які можуть бути використані для зниження рівня аварійності та підвищення безпеки дорожнього руху.

Зв'язок дослідницьких питань з першої лабораторної роботи з нинішніми:

Як особливості місцевості та дорожнього покриття (стан доріг, дорожні роботи тощо) впливають на ймовірність виникнення аварій та тяжкість?

Це питання пов'язане з:

- "Які дорожні умови частіше призводять до фатальних аварій?"
- "Чи відрізняється безпека дорожнього руху в сільській і міській місцевості?"

Як впливають характеристики автомобіля на кількість/серйозність аварій?

Це питання пов'язане з:

- "Чи пов'язаний вік автомобіля з ризиком фатальних аварій?"

Як впливає час/день/рік на кількість/серйозність аварій?

Це питання пов'язане з:

- "Чи відрізняється частота фатальних аварій у робочі та вихідні дні?"
- "Як змінюється кількість аварій залежно від сезону?"

Залежність фізичних особливостей водія на поведінку на дорозі?

Це питання пов'язане з:

- "Як впливає вік водія на ймовірність участі в аварії?"

Як фактори, на які не може впливати водій, впливають на кількість/серйозність аварій?

Це питання пов'язане з:

- "Чи пов'язаний вік автомобіля з ризиком фатальних аварій?"

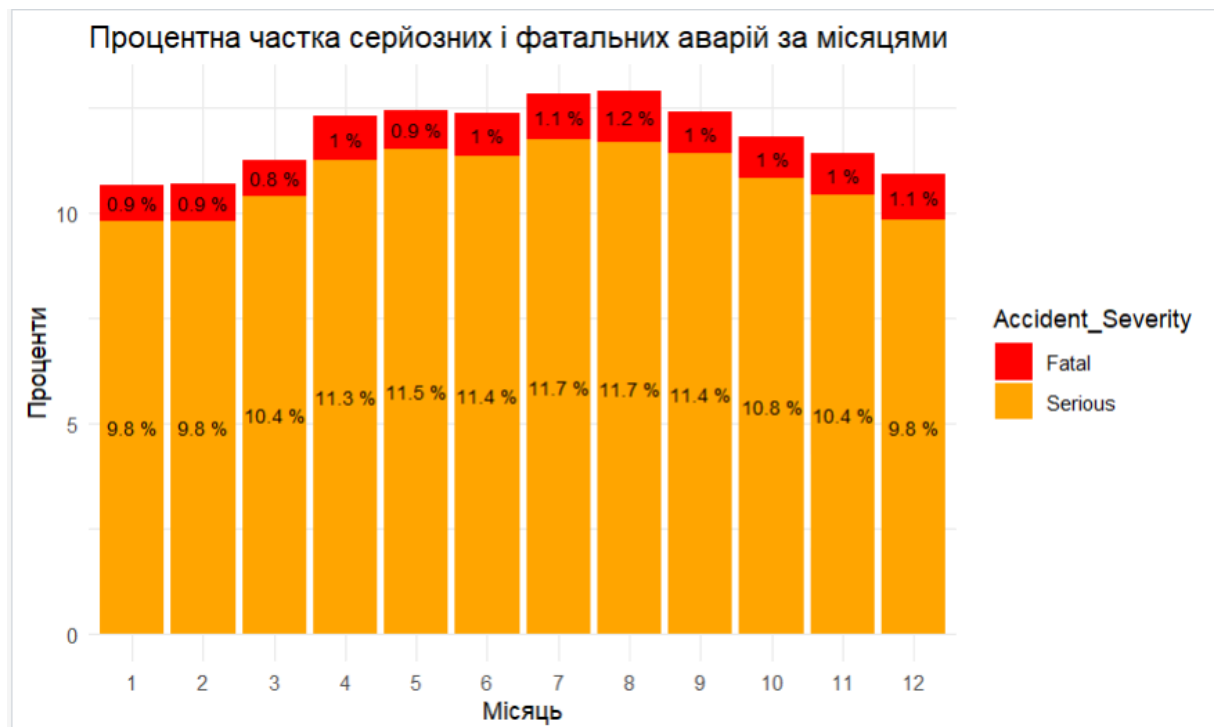
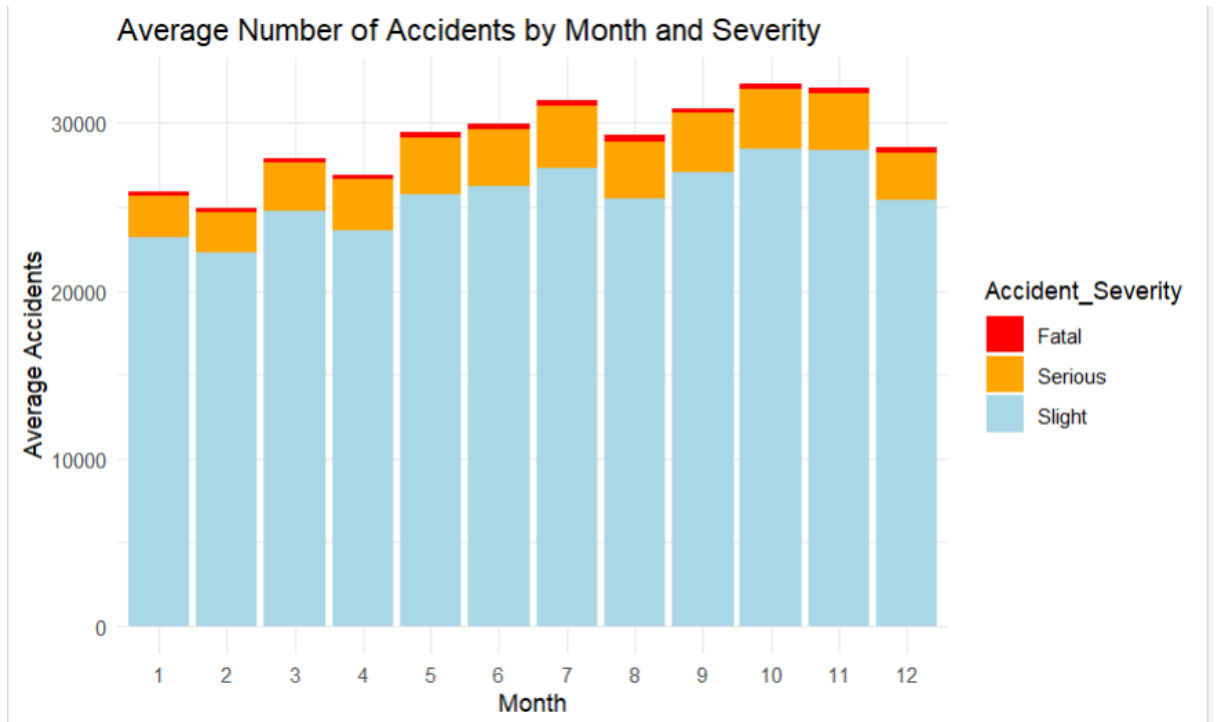
Гіпотези

1. Гіпотеза про рівність середньої кількості фатальних аварій в теплі та холодні місяці.
2. Гіпотеза про рівність середньої кількості фатальних аварій у робочі та вихідні дні.
3. Гіпотеза про рівність кількості фатальних аварій молодших та старших водіїв
4. Гіпотеза : Про рівність фатальних аварій автомобілей з більшим та меншим віком за роками
5. Гіпотеза про рівність середньої кількості за роками фатальних аварій у міській та сільській місцевості .

Обчислені статистики

Графіки з лабораторної 1

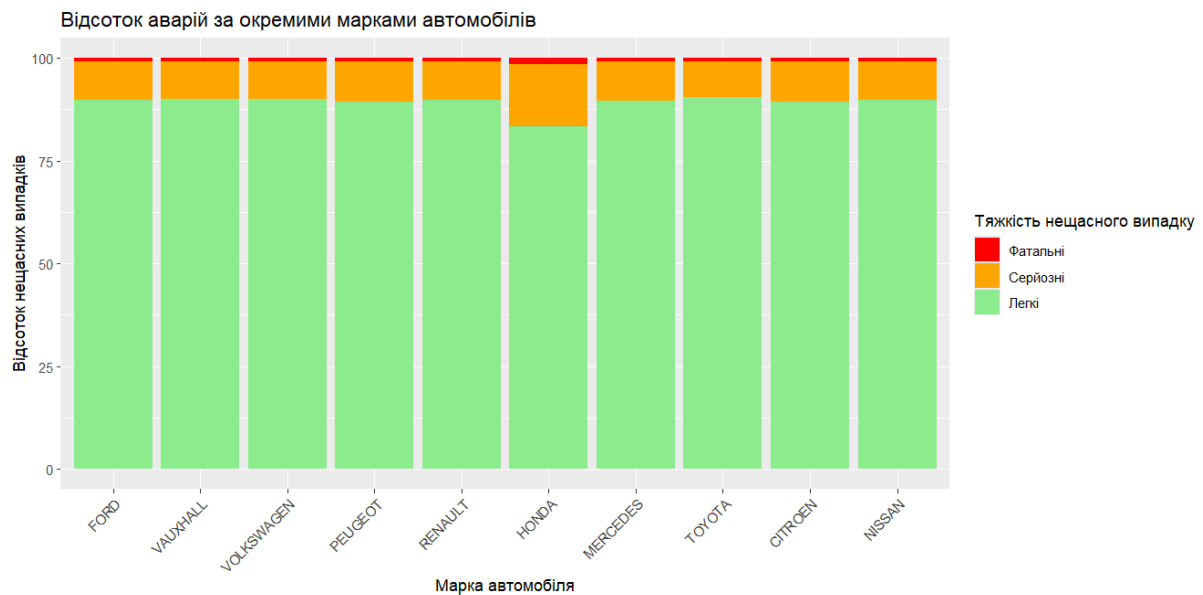
Спочатку були перероблені графіки з 1 лабораторної, та побудовані довірчі інтервали для них.



	Month	mean_serious	n	sd_serious	a	b
	<dbl>	<dbl>	<int>	<dbl>	<chr>	<chr>
1	1	254.2	2542	15.1255	253.61	254.79
2	2	244.3	2443	14.8280	243.71	244.89
3	3	290	2900	16.1555	289.41	290.59
4	4	302.6	3026	16.5027	302.01	303.19
5	5	339.1	3391	17.4697	338.51	339.69
6	6	340.2	3402	17.4980	339.61	340.79
7	7	368.1	3681	18.2014	367.51	368.69
8	8	341.3	3413	17.5263	340.71	341.89
9	9	352.9	3529	17.8216	352.31	353.49
10	10	349.3	3493	17.7305	348.71	349.89
11	11	334.2	3342	17.3430	333.61	334.79
12	12	281.2	2812	15.9085	280.61	281.79

	Month	mean_fatal	n	sd_fatal	a	b
	<dbl>	<dbl>	<int>	<dbl>	<dbl>	<dbl>
1	1	22.1	221	4.45982	21.5120	22.6880
2	2	21.9	219	4.43959	21.3120	22.4880
3	3	23.5	235	4.59891	22.9120	24.0880
4	4	27.7	277	4.99300	27.1120	28.2880
5	5	26.8	268	4.91121	26.2120	27.3880
6	6	29.9	299	5.18748	29.3120	30.4880
7	7	33.8	338	5.51543	33.2120	34.3880
8	8	35.7	357	5.66833	35.1120	36.2880
9	9	29.6	296	5.16140	29.0120	30.1880
10	10	31.7	317	5.34135	31.1120	32.2880
11	11	31.2	312	5.29906	30.6120	31.7880
12	12	30.6	306	5.24786	30.0120	31.1880

> |



	Car_Brand1	Accident_Severity	count	se	ci_lower	ci_upper
	<chr>	<chr>	<int>	<dbl>	<dbl>	<dbl>
1	FORD	Slight	36960	35.1	36891.	37029.
2	VAUXHALL	Slight	35979	34.6	35911.	36047.
3	VOLKSWAGEN	Slight	19709	25.6	19659.	19759.
4	PEUGEOT	Slight	18694	25.0	18645.	18743.
5	NULL	Slight	16380	23.4	16334.	16426.
6	RENAULT	Slight	15781	22.9	15736.	15826.
7	HONDA	Slight	14012	21.6	13970.	14054.
8	MERCEDES	Slight	13032	20.8	12991.	13073.
9	TOYOTA	Slight	12400	20.3	12360.	12440.
10	CITROEN	Slight	11422	19.5	11384.	11460.
11	FORD	Serious	3822	11.3	3800.	3844.
12	VAUXHALL	Serious	3615	11.0	3593.	3637.
13	HONDA	Serious	2560	9.24	2542.	2578.
14	NULL	Serious	2423	8.99	2405.	2441.
15	PEUGEOT	Serious	2081	8.33	2065.	2097.
16	VOLKSWAGEN	Serious	2025	8.22	2009.	2041.
17	RENAULT	Serious	1657	7.43	1642.	1672.
18	MERCEDES	Serious	1375	6.77	1362.	1388.
19	CITROEN	Serious	1252	6.46	1239.	1265.
20	TOYOTA	Serious	1206	6.34	1194.	1218.
21	FORD	Fatal	344	3.39	337.	351.
22	VAUXHALL	Fatal	322	3.28	316.	328.
23	HONDA	Fatal	240	2.83	234.	246.
24	VOLKSWAGEN	Fatal	162	2.32	157.	167.
25	PEUGEOT	Fatal	161	2.32	156.	166.
26	RENAULT	Fatal	149	2.23	145.	153.
27	MERCEDES	Fatal	128	2.07	124.	132.
28	CITROEN	Fatal	117	1.97	113.	121.
29	NULL	Fatal	107	1.89	103.	111.
30	TOYOTA	Fatal	100	1.83	96.4	104.

Загальна тенденція:

Серед марок автомобілів, Ford та Vauxhall мають найвищу загальну частоту аварій у всіх категоріях серйозності (легка, серйозна, смертельна). Це може вказувати на популярність цих марок автомобілів у використанні або на вищий ризик аварійності для цих моделей.

Довірчі інтервали:

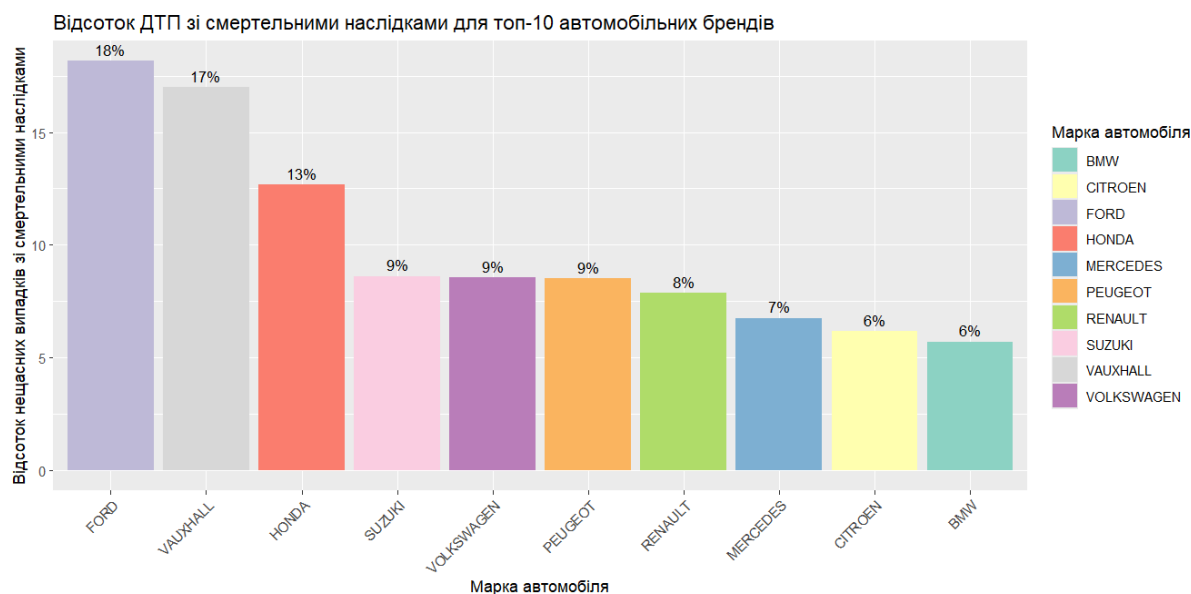
Вузькі довірчі інтервали: Для більшості марок із великою кількістю аварій довірчі інтервали є відносно вузькими, що свідчить про високу точність оцінок.

Легкі аварії: Більшість аварій має легкий характер, із значно більшими кількостями випадків порівняно з серйозними та смертельними. Це вказує на те, що хоча аварії з цими марками автомобілів часті, вони зазвичай не призводять до високої тяжкості травм.

Серйозні та смертельні аварії:

Показники серйозних та смертельних аварій нижчі, ніж легкі, але вони також вказують на певні тенденції. Марки автомобілів, як Ford, Vauxhall, і Honda, мають високі показники у всіх категоріях серйозності.

Марки Mercedes і Citroen мають порівняно нижчі частоти смертельних аварій, хоча їх присутність у категоріях серйозних і легких аварій залишається значною.



	Car_Brand1	count	se	ci_lower	ci_upper
	<chr>	<int>	<dbl>	<dbl>	<dbl>
1	FORD	344	18.5	308.	380.
2	VAUXHALL	322	17.9	287.	357.
3	HONDA	240	15.5	210.	270.
4	SUZUKI	163	12.8	138.	188.
5	VOLKSWAGEN	162	12.7	137.	187.
6	PEUGEOT	161	12.7	136.	186.
7	RENAULT	149	12.2	125.	173.
8	MERCEDES	128	11.3	106.	150.
9	CITROEN	117	10.8	95.8	138.
10	BMW	108	10.4	87.6	128.

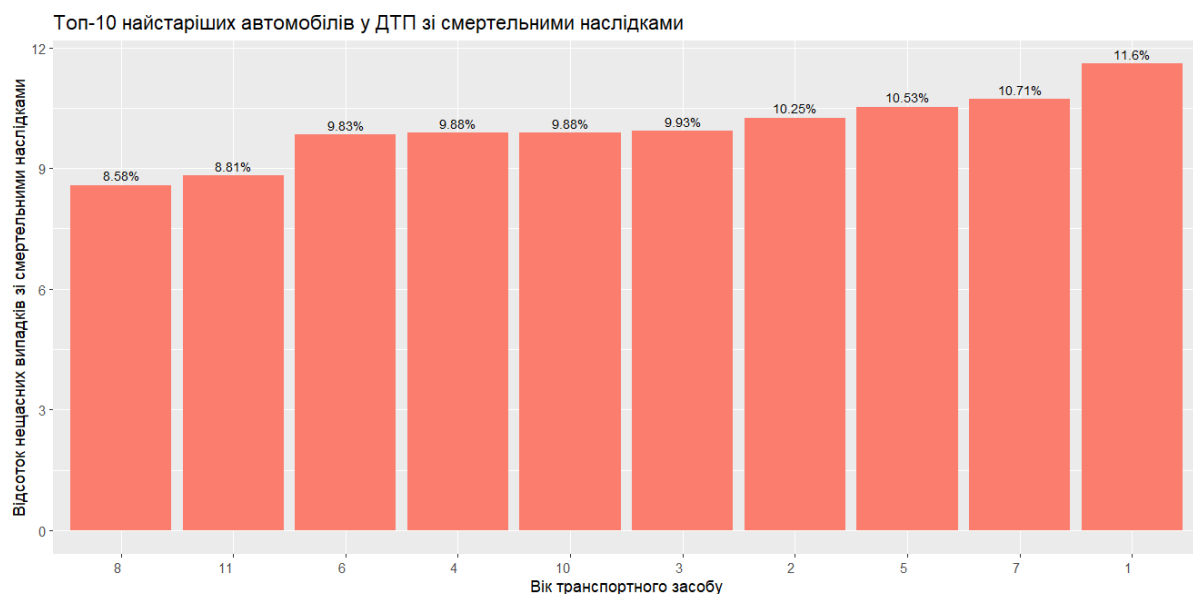
Розподіл смертельних аварій за марками:

Ford та Vauxhall лідирують за кількістю смертельних аварій. Це може бути пов'язано з високою популярністю цих марок в регіоні, де було зібрано дані, або з потенційними питаннями безпеки.

Ширина довірчих інтервалів:

Довірчі інтервали для більшості марок є порівняно вузькими, що свідчить про статистичну впевненість в отриманих оцінках. Наприклад, для Ford інтервал становить від 308 до 380, а для Vauxhall — від 287 до 357.

Вузькі довірчі інтервали для марок з великою кількістю смертельних аварій, як Ford і Vauxhall, підтверджують високу надійність даних.



Age_of_Vehicle	count	se	ci_lower	ci_upper	
	<i><int></i>	<i><int></i>	<i><dbl></i>	<i><dbl></i>	<i><dbl></i>
1	NA	457	21.4	415.	499.
2	1	250	15.8	219.	281.
3	7	231	15.2	201.	261.
4	5	227	15.1	197.	257.
5	2	221	14.9	192.	250.
6	3	214	14.6	185.	243.
7	4	213	14.6	184.	242.
8	10	213	14.6	184.	242.
9	6	212	14.6	183.	241.
10	11	190	13.8	163.	217.

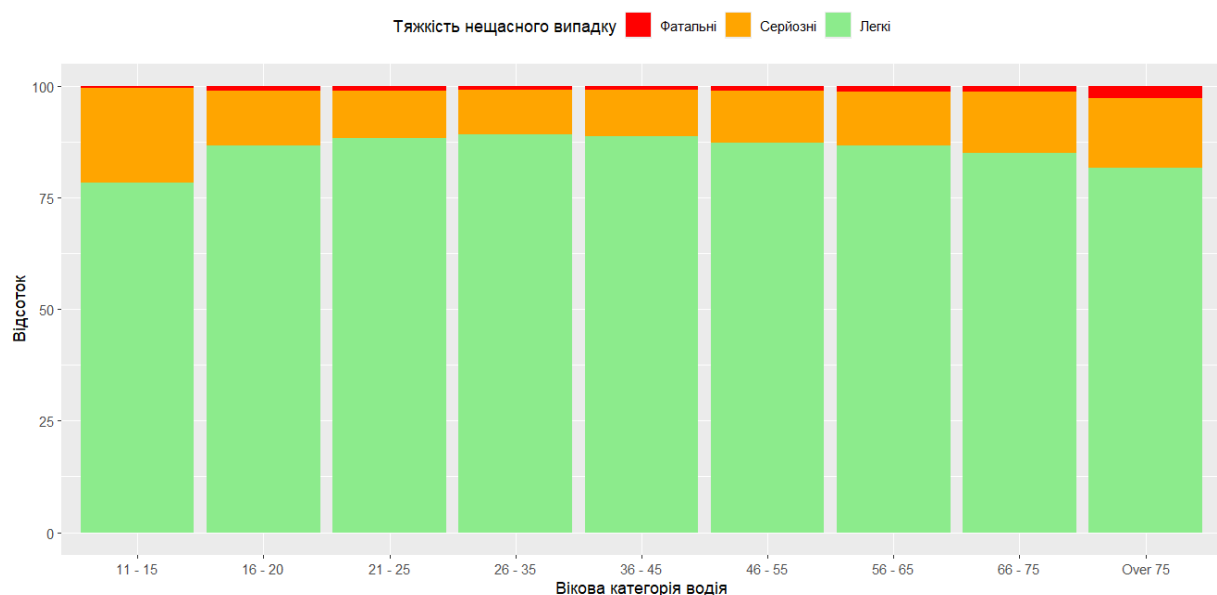
Вік Автомобілів:

Серед автомобілів з відомим віком, найбільше смертельних аварій стосується відносно нових автомобілів (1 рік) і автомобілів середнього віку (5-7 років). Це може свідчити про високий рівень активності водіїв цих автомобілів або про зниження безпеки автомобіля з часом.

Національний рівень смертельних аварій для автомобілів віком 10 і 11 років також високий, що може вказувати на потенційне зношування безпекових характеристик з віком автомобіля.

Довірчі інтервали:

Довірчі інтервали для всіх груп з віком автомобіля вузькі, що свідчить про велику точність оцінок кількості смертельних аварій в кожній категорії.



Age_Band_of_Driver1 <chr>	Accident_Severity <chr>	count <int>	se <dbl>	ci_lower <dbl>	ci_upper <dbl>
1 11 - 15	Fatal	2	0.239	1.53	2.47
2 11 - 15	Serious	130	1.93	126.	134.
3 11 - 15	Slight	475	3.68	468.	482.
4 16 - 20	Fatal	302	2.94	296.	308.
5 16 - 20	Serious	3840	10.5	3819.	3861.
6 16 - 20	Slight	26799	27.7	26745.	26853.
7 21 - 25	Fatal	402	3.39	395.	409.
8 21 - 25	Serious	4486	11.3	4464.	4508.
9 21 - 25	Slight	37375	32.7	37311.	37439.
10 26 - 35	Fatal	646	4.30	638.	654.
11 26 - 35	Serious	7386	14.5	7358.	7414.
12 26 - 35	Slight	66286	43.5	66201.	66371.
13 36 - 45	Fatal	589	4.10	581.	597.
14 36 - 45	Serious	6871	14.0	6844.	6898.
15 36 - 45	Slight	58981	41.1	58901.	59061.
16 46 - 55	Fatal	560	4	552.	568.
17 46 - 55	Serious	6079	13.2	6053.	6105.
18 46 - 55	Slight	45966	36.2	45895.	46037.
19 56 - 65	Fatal	374	3.27	368.	380.
20 56 - 65	Serious	3772	10.4	3752.	3792.
21 56 - 65	Slight	26777	27.7	26723.	26831.
22 66 - 75	Fatal	203	2.41	198.	208.
23 66 - 75	Serious	2198	7.92	2182.	2214.
24 66 - 75	Slight	13550	19.7	13511.	13589.
25 Over 75	Fatal	335	3.09	329.	341.
26 Over 75	Serious	1845	7.26	1831.	1859.
27 Over 75	Slight	9756	16.7	9723.	9789.

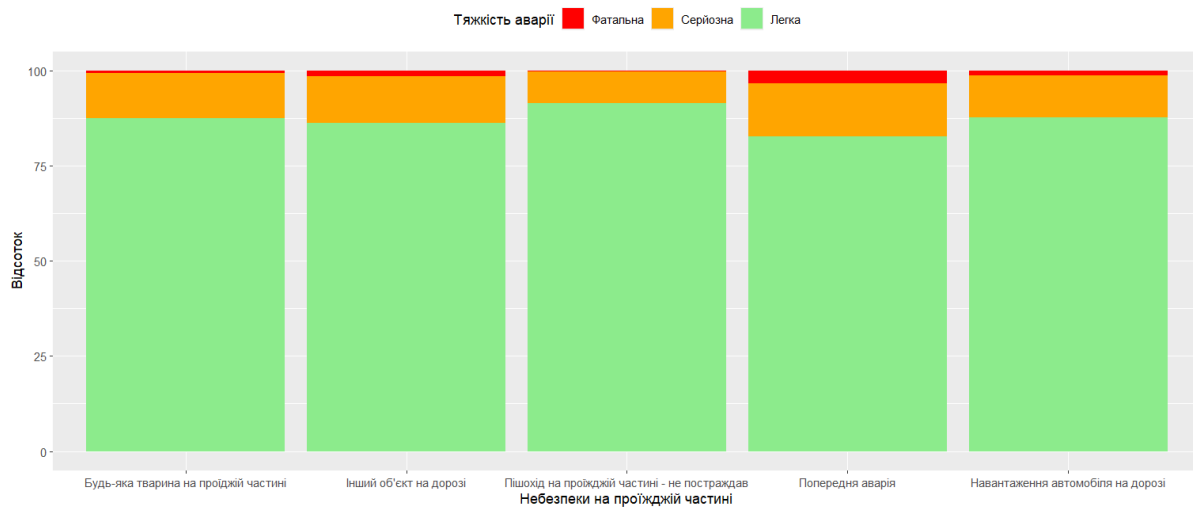
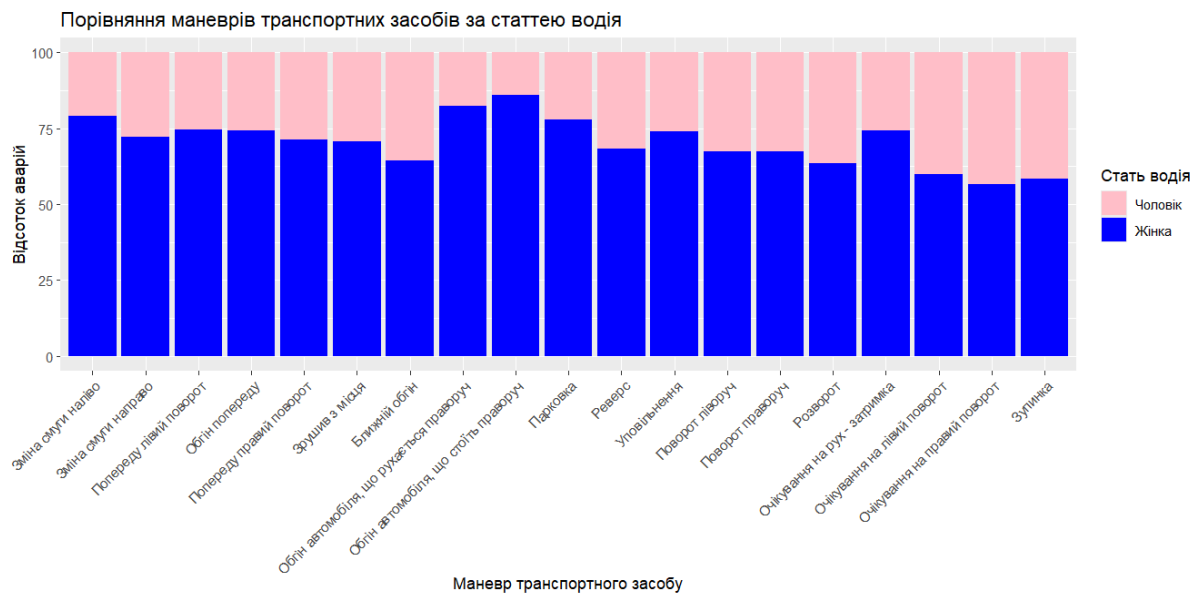
Зростання серйозності і частоти аварій з віком:

Вікові групи 16-20 років показують високу кількість серйозних (3840) та легких (26799) аварій. Це може вказувати на високий ризик участі в дорожньо-транспортних пригодах серед молодших водіїв.

Старші вікові групи (26-35, 36-45, 46-55) також мають високу частоту аварій, але їхні довірчі інтервали зазвичай вужчі, що свідчить про більшу стабільність оцінок в цих групах.

Висока серйозність аварій серед старших вікових груп:

Значення серйозних і фатальних аварій значно зростає серед вікових груп від 66 років і старше, зі збільшенням як середнього, так і верхнього значення довірчих інтервалів.



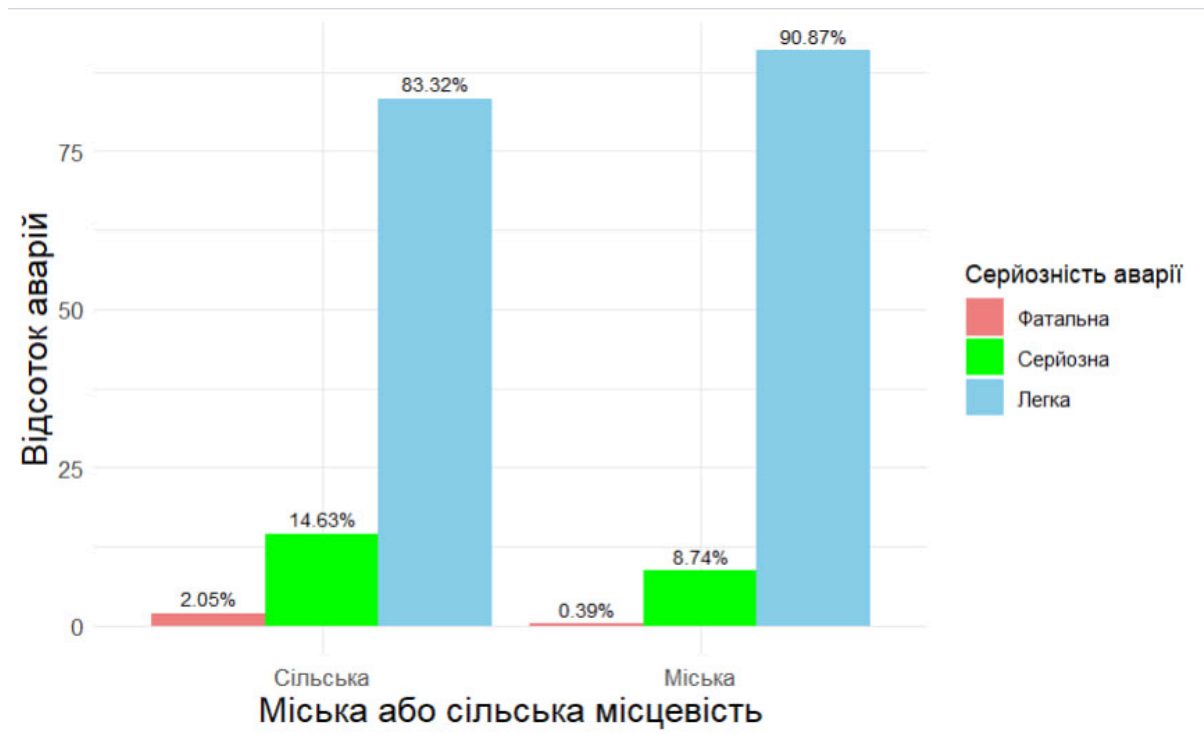
Carriageway_Hazards <chr>	Accident_Severity <chr>	count <int>	lower <dbl>	upper <dbl>
1 Any animal in carriageway (except ...	Fatal	5	6.17e-1	9.38e0
2 Any animal in carriageway (except ...	Serious	91	7.23e+1	1.10e2
3 Any animal in carriageway (except ...	Slight	679	6.28e+2	7.30e2
4 Data missing or out of range	Fatal	1	0	2.96e0
5 Data missing or out of range	Serious	3	0	6.39e0
6 Data missing or out of range	Slight	65	4.92e+1	8.08e1
7 Other object on road	Fatal	25	1.52e+1	3.48e1
8 Other object on road	Serious	206	1.78e+2	2.34e2
9 Other object on road	Slight	1451	1.38e+3	1.53e3
10 Pedestrian in carriageway - not in...	Fatal	1	0	2.96e0
11 Pedestrian in carriageway - not in...	Serious	40	2.76e+1	5.24e1
12 Pedestrian in carriageway - not in...	Slight	445	4.04e+2	4.86e2
13 Previous accident	Fatal	15	7.41e+0	2.26e1
14 Previous accident	Serious	62	4.66e+1	7.74e1
15 Previous accident	Slight	370	3.32e+2	4.08e2
16 Vehicle load on road	Fatal	5	6.17e-1	9.38e0
17 Vehicle load on road	Serious	47	3.36e+1	6.04e1
18 Vehicle load on road	Slight	368	3.30e+2	4.06e2

Вплив перешкоди:

Тварини на дорозі: Спостерігається значна кількість легких аварій (679 випадків), з порівняно високим довірчим інтервалом від 628 до 730. Це свідчить про те, що тварини на дорозі частіше призводять до менш серйозних аварій.

Інші об'єкти на дорозі: Висока кількість легких (1451) і серйозних (206) аварій, що може вказувати на ризики, пов'язані з цим типом перешкод. Довірчі інтервали також великі, що підкреслює частоту та серйозність цих інцидентів.

Попередня аварія на дорозі: Велика кількість легких (370) аварій із серйозними випадками (62) може вказувати на потенційну небезпеку затримки очищення місця попередніх аварій.



Довірчі інтервали для:

Кількість аварій і їх серйозність у міській і сільській місцевості.

Довірчі інтервали для серйозності

```
[1] "Довірчий інтервал для Slight"
      mean      sd      n      a      b
1 307992 191.0738 349411 307991.4 307992.6

[1] "Довірчий інтервал для Serious"
      mean      sd      n      a      b
1 37974 183.9755 349411 37973.39 37974.61

[1] "Довірчий інтервал для Fatal"
      mean      sd      n      a      b
1 3445 58.40406 349411 3444.806 3445.194
```

Можемо зробити висновки:

1. Серйозність аварій:

- Довірчий інтервал для "Slight" (легких) аварій має велику ширину, що свідчить про значну варіативність кількості таких подій.

- Для "Serious" (серйозних) аварій ширина довірчого інтервалу також значна, що свідчить про важливість вжиття заходів щодо попередження та реагування на ці події.

- Для "Fatal" (смертельних) аварій довірчий інтервал зазвичай вузький, що може вказувати на меншу варіативність кількості смертельних аварій порівняно з іншими категоріями.

Окремо для сільської місцевостей середні значення для категорій тяжкості ДТП:

```
> print(fatal_by_year_urban)
# A tibble: 1 x 5
      n mean_fatal sd_fatal      a      b
  <int>   <dbl>   <dbl>   <dbl>   <dbl>
1   865     86.5  8.82326 85.9120 87.0880 - Фатальні Місто

      n mean_serious sd_serious      a      b
  <int>   <dbl>   <dbl>   <dbl>   <dbl>
1 19524    1952.4  41.9185 1951.81 1952.99 - Серйозні Місто

      n mean_slight sd_slight      a      b
  <int>   <dbl>   <dbl>   <dbl>   <dbl>
1 202916    20291.6 135.139 20291.0 20292.2 - Легкі Місто

      n mean_fatal sd_fatal      a      b
  <int>   <dbl>   <dbl>   <dbl>   <dbl>
1   2580     258  15.2381 257.412 258.588 - Фатальні Село

      n mean_serious sd_serious      a      b
  <int>   <dbl>   <dbl>   <dbl>   <dbl>
1 18450    1845  40.7492 1844.41 1845.59 - Серйозні Село

      n mean_slight sd_slight      a      b
  <int>   <dbl>   <dbl>   <dbl>   <dbl>
1 105075    10507.5  97.2458 10506.9 10508.1 - Легкі село
```

Довірчі інтервали для типу місцевості

```
[1] "Довірчий інтервал для Urban"
      mean      sd      n      a      b
1 223305 283.8892 349411 223304.1 223305.9
[1] "Довірчий інтервал для Rural"
      mean      sd      n      a      b
1 126105 283.8887 349411 126104.1 126105.9
[1] "Довірчий інтервал для Unallocated"
      mean      sd      n      a      b
1      1 0.9999986 349411 0.9966843 1.003316
```

Можемо зробити висновки:

1. Міська місцевість (Urban): Середня кількість аварій у міській місцевості є вищою, ніж у сільській місцевості.

2. Сільська місцевість (Rural): Хоча середня кількість аварій у сільській місцевості менша, ніж у міській.
3. Нерозподілені (Unallocated): Середня кількість аварій у цій категорії є дуже низькою, що може вказувати на відсутність чіткого визначення місця події або на технічні проблеми з відображенням даних. Ці дані можуть бути виключеними або потребують подальшого аналізу та уточнення.

Погодні умови під час ДТП

Довірчі інтервали для погодних умов

```
[1] "Довірчий інтервал для Fine no high winds"
      mean      sd      n      a      b
1 283155 231.7162 349411 283154.2 283155.8
[1] "Довірчий інтервал для Snowing no high winds"
      mean      sd      n      a      b
1 2321 48.01648 349411 2320.841 2321.159
[1] "Довірчий інтервал для Raining no high winds"
      mean      sd      n      a      b
1 40284 188.7846 349411 40283.37 40284.63
[1] "Довірчий інтервал для Other"
      mean      sd      n      a      b
1 6604 80.49337 349411 6603.733 6604.267
[1] "Довірчий інтервал для Fine + high winds"
      mean      sd      n      a      b
1 4383 65.78769 349411 4382.782 4383.218
[1] "Довірчий інтервал для Fog or mist"
      mean      sd      n      a      b
1 1623 40.1928 349411 1622.867 1623.133
[1] "Довірчий інтервал для Raining + high winds"
      mean      sd      n      a      b
1 4617 67.49809 349411 4616.776 4617.224
[1] "Довірчий інтервал для Unknown"
      mean      sd      n      a      b
1 5953 76.4956 349411 5952.746 5953.254
[1] "Довірчий інтервал для Snowing + high winds"
      mean      sd      n      a      b
1 444 21.05792 349411 443.9302 444.0698
[1] "Довірчий інтервал для Data missing or out of range"
      mean      sd      n      a      b
1 27 5.195952 349411 26.98277 27.01723
```

На основі наданих довірчих інтервалів для погодних умов під час ДТП можна зробити такі висновки щодо їхньої довжини: загальна тенденція полягає в тому, що більшість довірчих інтервалів є вузькими, що свідчить про відносну точність оцінок середньої кількості аварій у різних погодних

умовах. Однак варто відзначити, що деякі категорії можуть мати більшу неоднорідність даних, що відображається в більших довірчих інтервалах.

Залежність відсотка кількості аварій від світлових умов

Довірчі інтервали для світлових умов

```
[1] "Довірчий інтервал для Daylight"
      mean      sd      n      a      b
1 267835 250.0614 349411 267834.2 267835.8
[1] "Довірчий інтервал для Darkness - lights lit"
      mean      sd      n      a      b
1 63986 228.6231 349411 63985.24 63986.76
[1] "Довірчий інтервал для Darkness - lighting unknown"
      mean      sd      n      a      b
1 3900 62.10048 349411 3899.794 3900.206
[1] "Довірчий інтервал для Darkness - lights unlit"
      mean      sd      n      a      b
1 1453 38.0389 349411 1452.874 1453.126
[1] "Довірчий інтервал для Darkness - no lighting"
      mean      sd      n      a      b
1 12237 108.6666 349411 12236.64 12237.36
```

Отже, загальна тенденція полягає в тому, що більшість довірчих інтервалів є вузькими, що свідчить про відносну точність оцінок кількості аварій у різних світлових умовах. Однак варто відзначити, що категорія "Darkness - no lighting" має більшу неоднорідність даних, що відображається в більшому довірчому інтервалі.

Кількість аварій за станом дорожнього покриття та серйозністю

Довірчі інтервали для стану дорожнього покриття

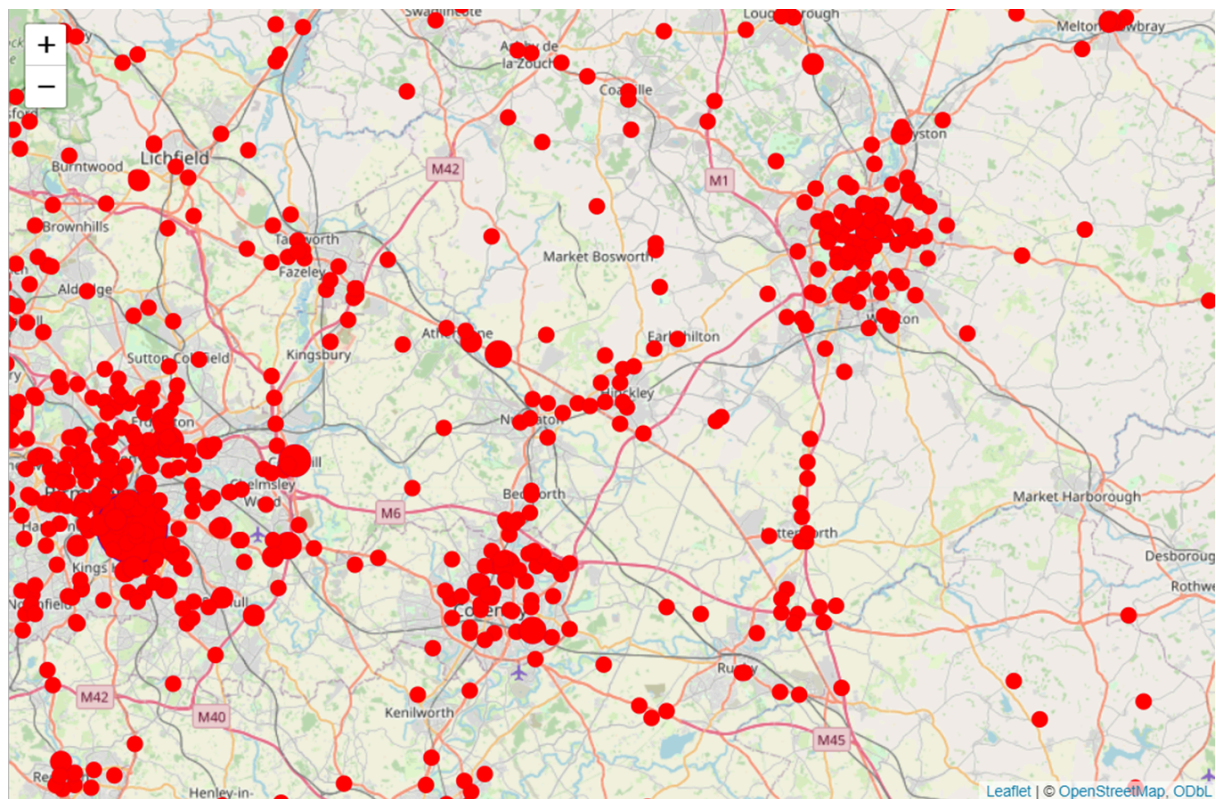
```
[1] "Довірчий інтервал для Dry"
      mean      sd      n      a      b
1 246425 269.5027 349411 246424.1 246425.9
[1] "Довірчий інтервал для Wet or damp"
      mean      sd      n      a      b
1 94438 262.5138 349411 94437.13 94438.87
[1] "Довірчий інтервал для Frost or ice"
      mean      sd      n      a      b
1 5354 72.60827 349411 5353.759 5354.241
[1] "Довірчий інтервал для Flood over 3cm. deep"
      mean      sd      n      a      b
1 348 18.64547 349411 347.9382 348.0618
[1] "Довірчий інтервал для Data missing or out of range"
      mean      sd      n      a      b
1 508 22.52247 349411 507.9253 508.0747
```

```
[1] "Довірчий інтервал для Snow"
```

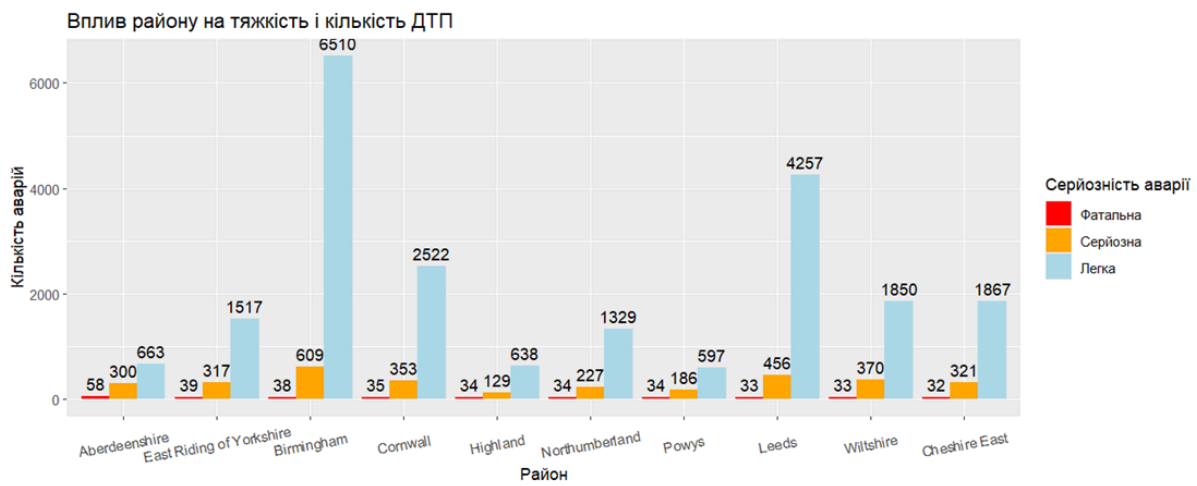
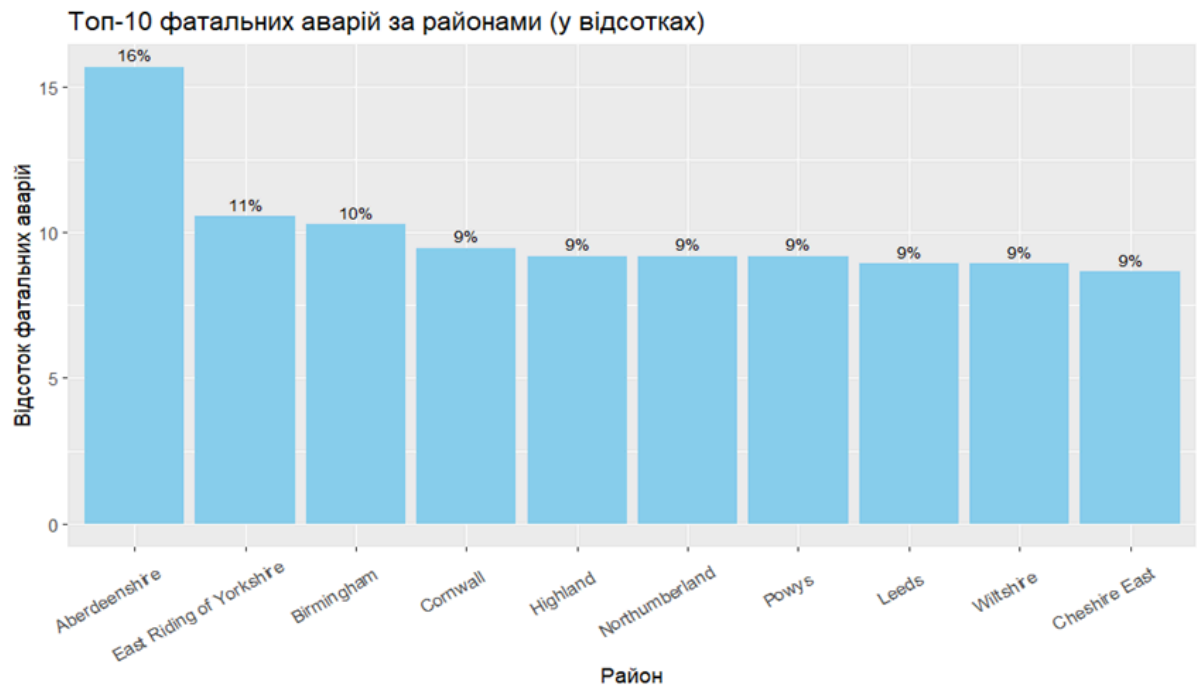
	mean	sd	n	a	b
1	2338	48.19083	349411	2337.84	2338.16

Отже, загальна тенденція полягає в тому, що більшість довірчих інтервалів є вузькими, що свідчить про відносну точність оцінок кількості аварій у різних умовах дорожнього покриття. Однак варто відзначити, що деякі категорії мають більшу неоднорідність даних, що відображається в більших довірчих інтервалах.

Мапа аварій:



Відображення кількості аварій за районами.

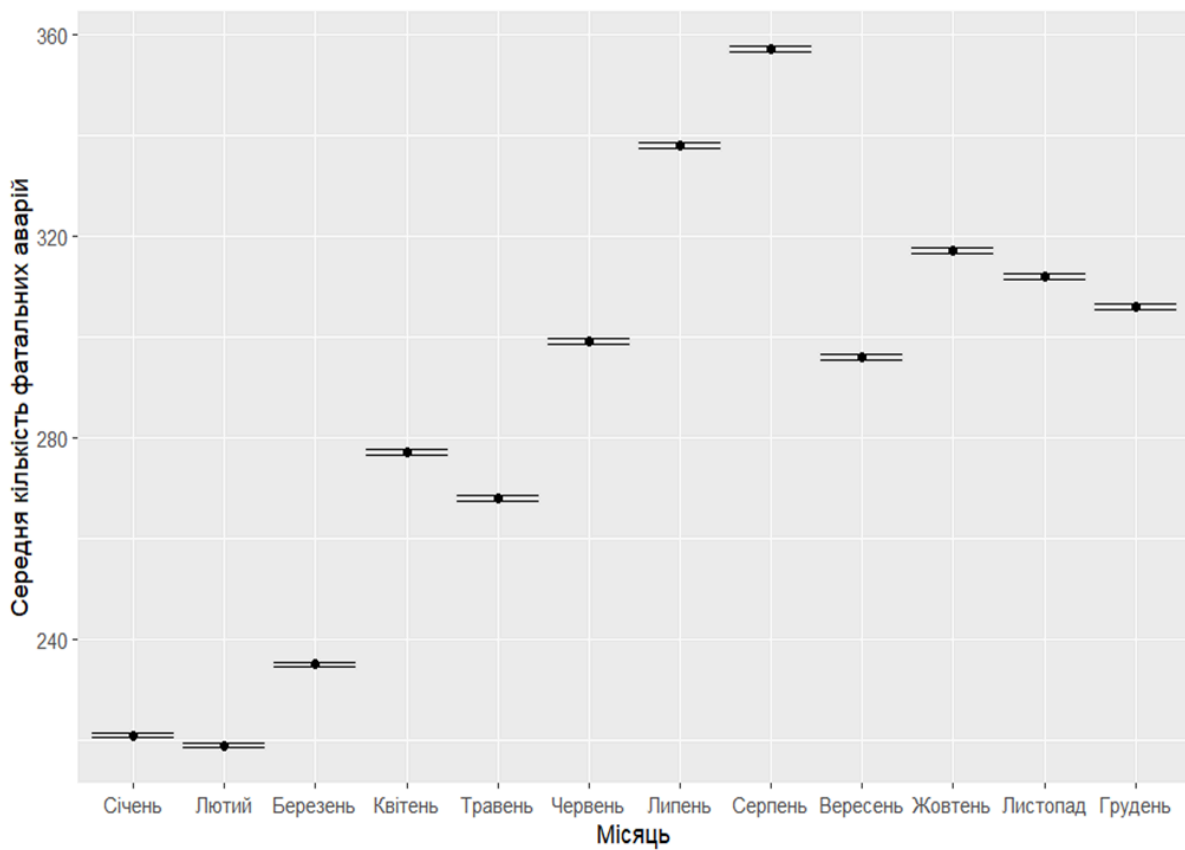
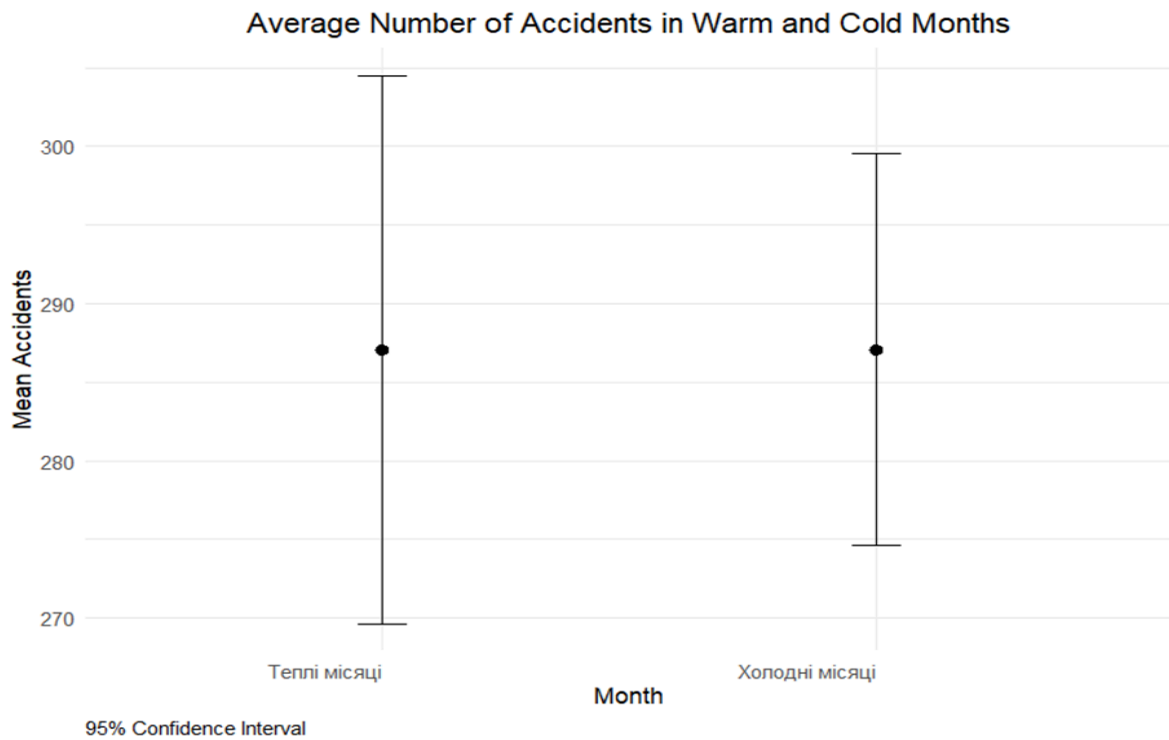


Гіпотеза 1. Про рівність середньої кількості фатальних аварій в теплі та холодні місяці.

```
[1] "Довірчий інтервал для 1"
  mean      sd      n      a      b
1  161 12.26476 2451 160.5144 161.4856
[1] "Довірчий інтервал для 2"
  mean      sd      n      a      b
1  170 12.57811 2451 169.502 170.498
[1] "Довірчий інтервал для 3"
  mean      sd      n      a      b
1  181 12.94734 2451 180.4874 181.5126
[1] "Довірчий інтервал для 4"
  mean      sd      n      a      b
1  203 13.64503 2451 202.4598 203.5402
[1] "Довірчий інтервал для 5"
  mean      sd      n      a      b
1  192 13.30262 2451 191.4734 192.5266
[1] "Довірчий інтервал для 6"
  mean      sd      n      a      b
1  218 14.09292 2451 217.4421 218.5579
[1] "Довірчий інтервал для 7"
  mean      sd      n      a      b
1  238 14.65911 2451 237.4197 238.5803
[1] "Довірчий інтервал для 8"
  mean      sd      n      a      b
1  249 14.95673 2451 248.4079 249.5921
[1] "Довірчий інтервал для 9"
  mean      sd      n      a      b
1  206 13.73631 2451 205.4562 206.5438
[1] "Довірчий інтервал для 10"
  mean      sd      n      a      b
1  208 13.79668 2451 207.4538 208.5462
[1] "Довірчий інтервал для 11"
  mean      sd      n      a      b
1  213 13.94595 2451 212.4479 213.5521
[1] "Довірчий інтервал для 12"
  mean      sd      n      a      b
1  212 13.91629 2451 211.4491 212.5509

# A tibble: 1 x 7
  mean      sd      n      a      b
<dbl> <dbl> <int> <dbl> <dbl>
1  287.   44.5      7  275.   300.

# A tibble: 1 x 7
  mean      sd      n      a      b
<dbl> <dbl> <int> <dbl> <dbl>
1  287.   44.5      5  270.   305.
```



Аналіз даних про кількість аварій за місяцями та ступенем серйозності аварій(фатальних) вказує на кілька ключових тенденції на дорозі:

Розподіл серйозності аварій у холодні та теплі місяці року:

Як бачимо на графіку, кількість смертельних аварій у теплі місяці вища, ніж у холодні. Однією з причин, може стати збільшений рух на дорогах через кращі погодні умови.

Розподіл по місяцям:

Серпень і липень: Для цих місяців довірчі інтервали для середнього значення є дуже близькими, що може свідчити про те, що середня кількість чи не суттєво відрізняються в ці місяці, але залишаються найвищими за весь час.

Червень і липень та Лютий: Відносно великі значення SD у літні місяці можуть вказувати на значний розкид даних і можливість значних коливань в досліджуваних параметрах протягом цих місяців, у порівнянні з холодним місяцем.

Грудень і квітень: Спостереження для цих місяців також мають схожі довірчі інтервали, але нижні межі довірчих інтервалів для середніх значень грудня нижчі, що може вказує на зниження показників протягом зимових місяців.

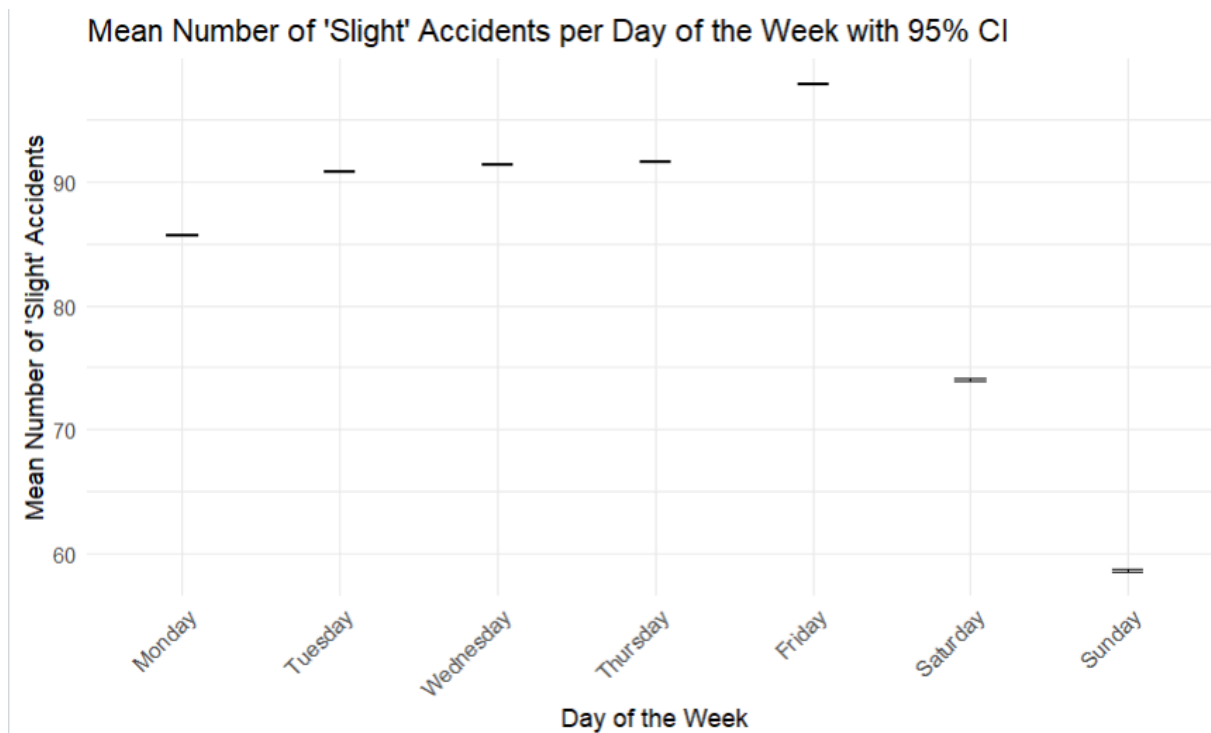
Березень і квітень: Також відносно близькі значення довірчих інтервалів, що може вказувати на подібність в середніх значеннях параметрів для цих місяців.

Гіпотеза 2. Про рівність середньої кількості фатальних аварій у робочі та вихідні дні.

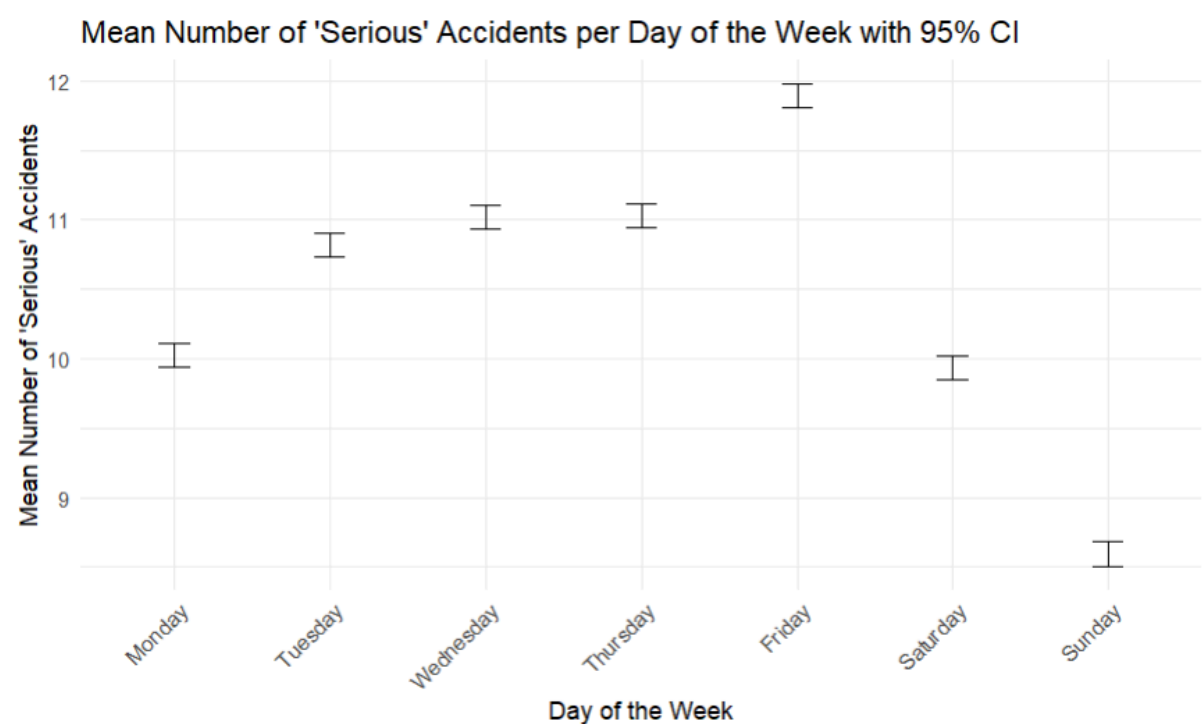
Довірчі інтервали для середньої кількості легких, серйозних та фатальних аварій по дням тижня, а також графіки для візуалізації:

	Day_of_Week	mean_slight	sd_slight	n	a	b
	<chr>	<dbl>	<dbl>	<int>	<dbl>	<dbl>
1	Monday	85.7356	8.46487	44754	85.6572	85.8141
2	Tuesday	90.8317	8.66352	47505	90.7538	90.9096
3	Wednesday	91.4195	8.68383	47721	91.3416	91.4975
4	Thursday	91.6628	8.69292	47848	91.5849	91.7407
5	Friday	97.9098	8.91683	51011	97.8324	97.9872
6	Saturday	73.9981	7.96794	38553	73.9185	74.0776
7	Sunday	58.6207	7.21371	30600	58.5399	58.7015

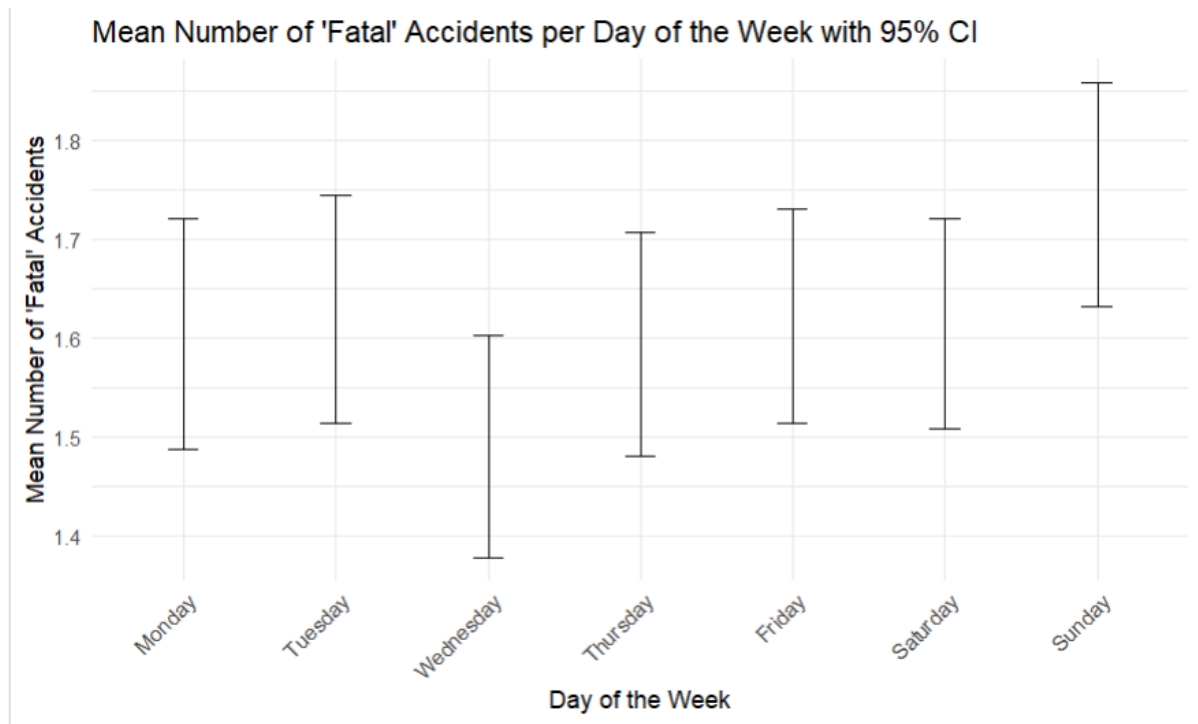
>



	Day_of_Week	mean_serious	sd_serious	n	a	b
	<chr>	<dbl>	<dbl>	<int>	<dbl>	<dbl>
1	Monday	10.0211	3.13502	5221	9.93608	10.1062
2	Tuesday	10.8205	3.25491	5605	10.7353	10.9057
3	Wednesday	11.0173	3.28382	5718	10.9322	11.1025
4	Thursday	11.0289	3.28550	5724	10.9438	11.1140
5	Friday	11.8942	3.40913	6185	11.8093	11.9792
6	Saturday	9.93629	3.12181	5147	9.85101	10.0216
7	Sunday	8.59332	2.90659	4374	8.50718	8.67946

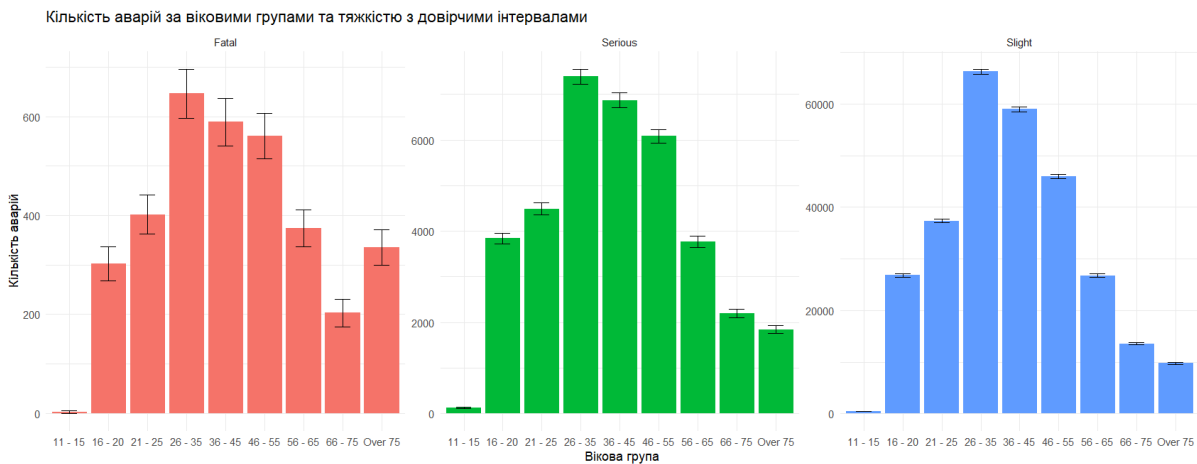
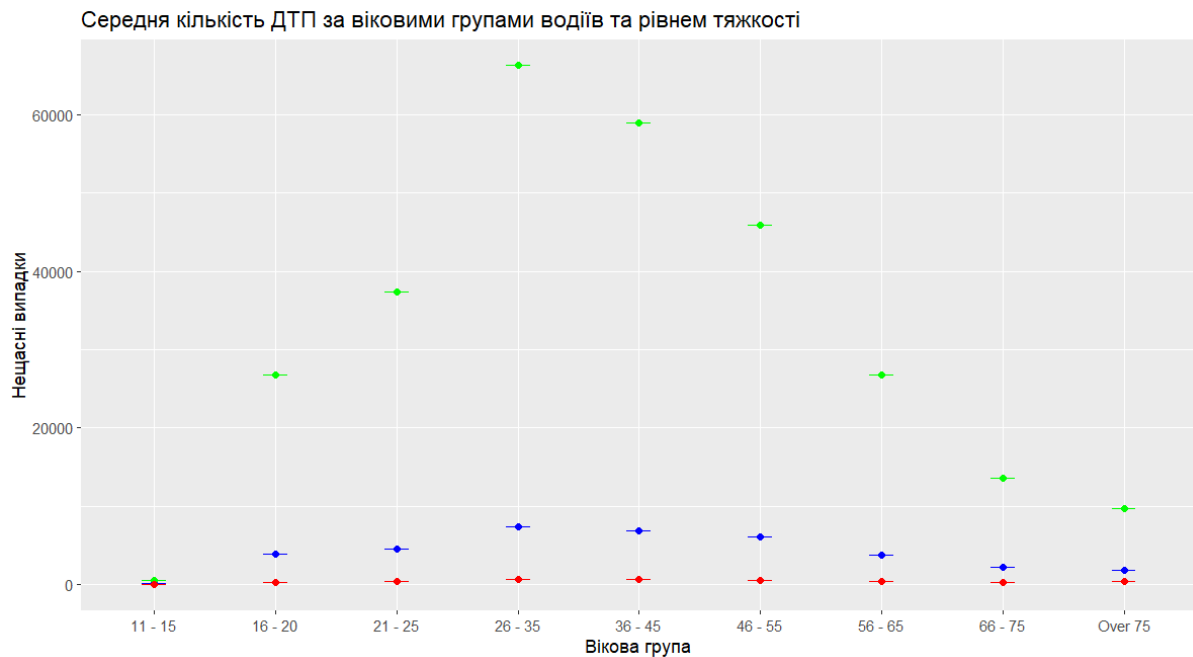


```
# A tibble: 7 × 6
  Day_of_Week mean_fatal sd_fatal   n      a      b
  <chr>         <dbl>   <dbl> <int>   <dbl>   <dbl>
1 Monday       1.60357 1.26269  449 1.48678 1.72037
2 Tuesday      1.62847 1.27250  469 1.51331 1.74364
3 Wednesday    1.49013 1.21771  453 1.37800 1.60227
4 Thursday     1.59333 1.25892  478 1.48048 1.70619
5 Friday       1.62154 1.27022  527 1.51309 1.72999
6 Saturday     1.61471 1.26769  549 1.50866 1.72075
7 Sunday       1.74497 1.31710  520 1.63176 1.85817
```



Як можна побачити з графіків, найбільше легких аварій припадає на будні дні, а надто на п'ятницю, а в суботу і неділю їх кількість набагато нижча. Приблизно аналогічна картина спостерігається і для серйозних аварій. Натомість, коли мова заходить про фатальні аварій ми можемо побачити деякі цікаві моменти: довірчі інтервали для всіх будніх днів приблизно однаковий, за виключенням середи, де значення значно нижчі від інших. І навпаки, у неділю, значення границь довірчого інтервалу і середнього значення вищі, ніж у інші дні.

Гіпотеза 3. Гіпотеза про рівність кількості фатальних аварій молодших та старших водіїв



Аналіз даних про кількість аварій за віковими групами водіїв та ступенем серйозності аварій вказує на кілька ключових тенденцій і можливостей для покращення безпеки на дорозі:

Розподіл серйозності аварій по вікових групах:

Як видно з даних, кількість легких аварій (Slight) значно перевищує кількість серйозних (Serious) та смертельних (Fatal) аварій у кожній віковій

групі. Це може свідчити про те, що більшість аварій не призводить до серйозних травм або смерті, що є позитивним аспектом з точки зору безпеки.

Пік аварійності в середньому віці:

Вікові групи 21-25, 26-35, та 36-45 років мають найвищі показники аварійності в категоріях серйозних і легких аварій. Це може вказувати на високий рівень дорожнього руху серед осіб цих вікових груп або на недостатній рівень обережності чи досвіду.

Смертельні аварії зосереджені у старших вікових групах:

Вікові групи "66 - 75" та "Over 75" мають відносно високий відсоток смертельних аварій порівняно з іншими групами. Це може свідчити про підвищені ризики для старших водіїв, можливо, через зниження фізичних або когнітивних здібностей.

Довірчі інтервали:

Вузькі довірчі інтервали для великих кількостей випадків (особливо в категорії Slight) підтверджують високу точність оцінок кількості аварій. Водночас, ширші довірчі інтервали у категорії Fatal в деяких молодших вікових групах (наприклад, "11 - 15") можуть вказувати на більшу статистичну невизначеність через меншу кількість даних.

Гіпотеза 4. Про рівність фатальних аварій автомобілей з більшим та меншим віком за роками

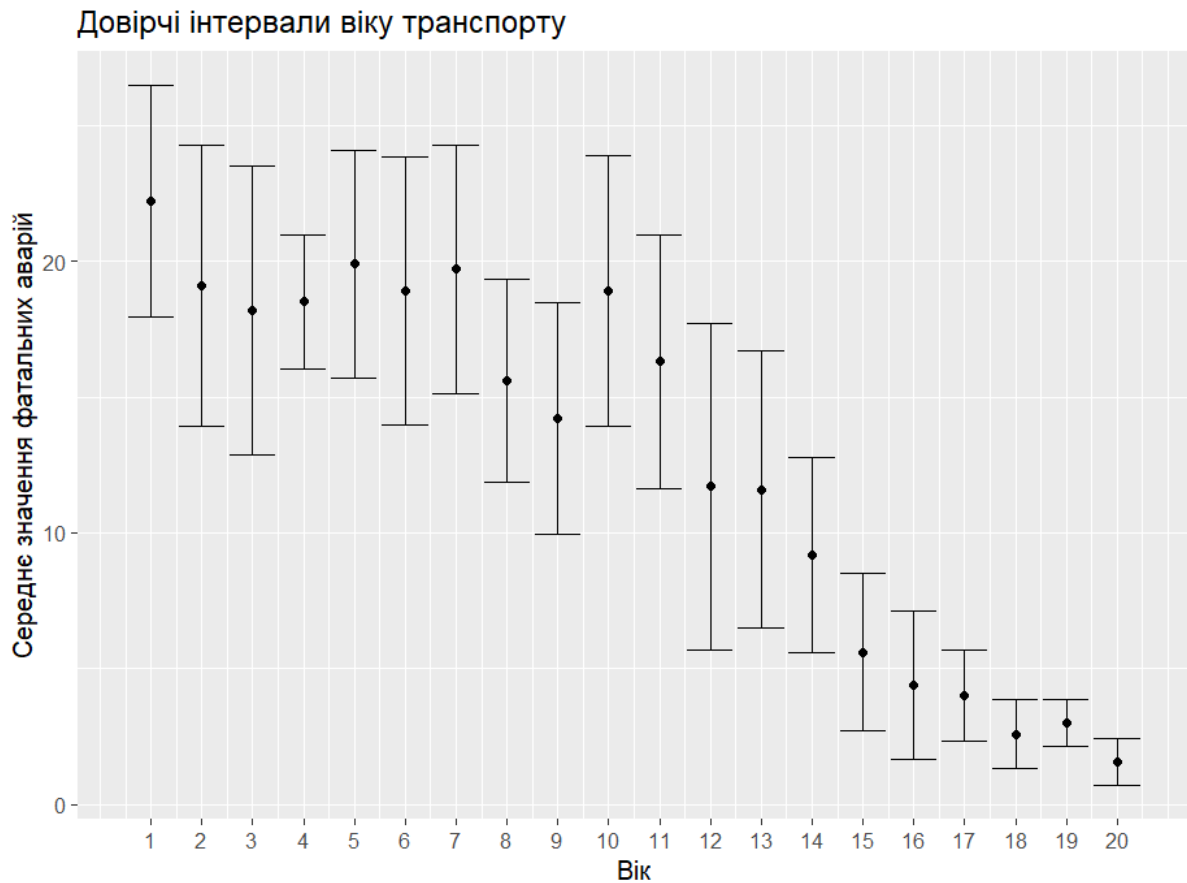
Медіана віку автомобілів у фатальних аваріях: 7
95% довірчий інтервал медіани: 6.733564 - 7.266436

Вік автомобілів і фатальні аварії: Медіана в 7 років вказує на те, що половина всіх автомобілів, що беруть участь у фатальних аваріях, молодша або дорівнює 7 рокам, а інша половина - старша. Це передбачає, що серед автомобілів, які беруть участь у фатальних аваріях, немає значного перекосу в бік дуже нових або дуже старих машин. Однак значна частина автомобілів, що беруть участь в аваріях, все ж таки має вік близько 7 років.

Тіснота довірчого інтервалу: Вузкий довірчий інтервал медіани показує, що середній вік автомобілів у фатальних аваріях досить стабільно зосереджений навколо 7 років на великому масиві даних. Це свідчить про низький ступінь невизначеності в оцінці медіанного віку автомобілів у таких аваріях.

Розглянемо довірчі інтервали для віку авто що менше 20

	mean	sd	n	a	b
	<dbl>	<dbl>	<int>	<dbl>	<dbl>
1	22.2	6.86	10	17.9	26.5
2	19.1	8.37	10	13.9	24.3
3	18.2	8.57	10	12.9	23.5
4	18.5	3.98	10	16.0	21.0
5	19.9	6.77	10	15.7	24.1
6	18.9	7.98	10	14.0	23.8
7	19.7	7.38	10	15.1	24.3
8	15.6	6.04	10	11.9	19.3
9	14.2	6.86	10	9.95	18.5
10	18.9	8.06	10	13.9	23.9
11	16.3	7.54	10	11.6	21.0
12	11.7	9.72	10	5.68	17.7
13	11.6	8.21	10	6.51	16.7
14	9.2	5.79	10	5.61	12.8
15	5.6	4.67	10	2.70	8.50
16	4.4	4.43	10	1.66	7.14
17	4	2.71	10	2.32	5.68
18	2.6	2.01	10	1.35	3.85
19	3	1	5	2.12	3.88
20	1.57	1.13	7	0.731	2.41

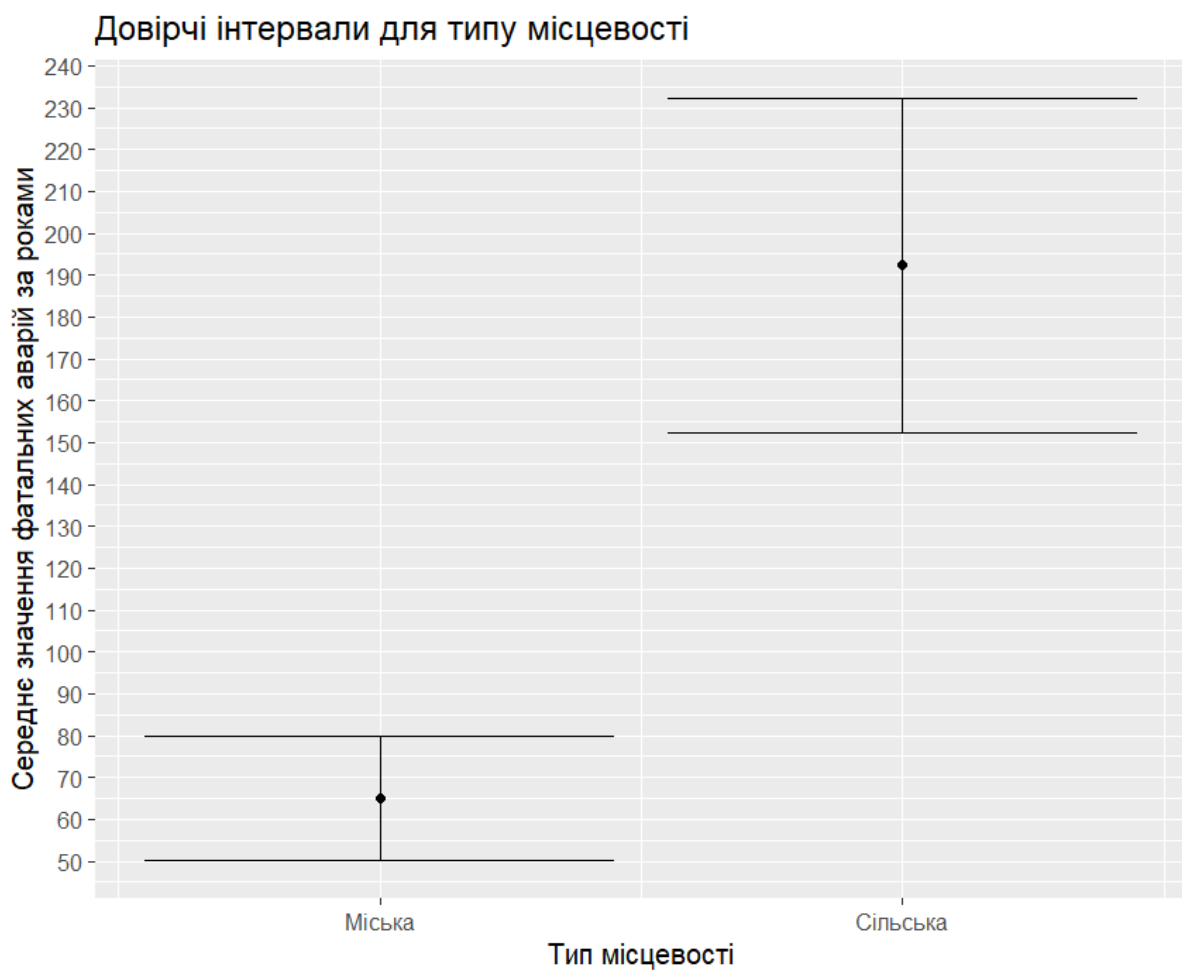


Висновок: оскільки довірчі інтервали перетинаються, можна припустити, про відсутність статистично значущої різниці в середній кількості фатальних аварій між цими групами. Це означає, що вплив віку автомобіля на рівень фатальних аварій, ймовірно, не є значущим.

Гіпотеза 5. Про рівність середньої кількості за роками фатальних аварій у міській та сільській місцевості .

Довірчі інтервали для міської і сільської місцевості відповідно

	mean	sd	n	a	b
	<dbl>	<dbl>	<int>	<dbl>	<dbl>
1	65	24.1	10	50.1	79.9
2	192.	64.4	10	152.	232.



Оскільки інтервали не перетинаються, можна припустити, що є статистично значуща різниця у середній кількості фатальних аварій між міською та сільською місцевостями. У цьому випадку, в середньому, в сільській місцевості кількість фатальних аварій є вищою, ніж у міській місцевості.

Протестовані гіпотези

Гіпотеза 1. Про рівність середньої кількості аварій в теплі та холодні місяці.

```
$mean_x  
[1] 220.6
```

```
$mean_y  
[1] 192.5714
```

```
$p_value  
[1] 0.06467324
```

```
$conf_int  
[1] -2.138961 58.196104
```

За результатами тесту Волда отримано наступні значення:

- Середнє значення для теплих місяців: 220.6
- Середнє значення для холодних місяців: 192.5714
- Р-значення: 0.06467324
- Довірчий інтервал для різниці середніх між теплими та холодними місяцями: від -2.138961 до 58.196104

З огляду на отримані результати можна зробити висновок, що на рівні значущості 0.05 немає статистично значущої різниці у середній кількості фатальних аварій між теплими та холодними місяцями. Р-значення (0.06467324) перевищує зазначений рівень значущості, що означає відсутність достовірних даних для відхилення нульової гіпотези про рівність середніх.

Welch Two Sample t-test

```
data: ci11$mean_accidents and ci22$mean_accidents
t = -2.1273, df = 8.3413, p-value = 0.06467
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -58.196104  2.138961
sample estimates:
mean of x mean of y
192.5714 220.6000
```

Результат Welch Two Sample t-test показує, що p-value дорівнює 0.06467, що більше типового рівня значимості 0.05. Це означає, що немає статистично значущих доказів на користь того, що середні значення кількості фатальних аварій в холодні та теплі місяці відрізняються. Однак, слід звернути увагу, що p-value досить близьке до рівня 0.05, тому можуть існувати деякі докази на користь різниці, хоча вони не досягають рівня статистичної значимості на звичайному рівні.

Отже, на підставі цього тесту немає достатніх підстав відкидати нульову гіпотезу про рівність середніх значень кількості фатальних аварій у теплі та холодні місяці.

```
Дисперсія для холодних місяців: 463.619047619048
Дисперсія для теплих місяців: 536.8
```

```
[1] "F-статистика: 1.15784716516023"
> print(paste("Ступені свободи:", df1, "та", df2))
[1] "Ступені свободи: 4 та 6"
> print(paste("p-value:", p_value))
[1] "p-value: 0.829359079021096"
```

Результати тесту Фішера показують, що p-value = 0.8294, що значно більше за звичайний рівень значущості 0.05. Отже, ми не маємо достатніх доказів для відкидання нульової гіпотези про рівність дисперсій між теплими та холодними місяцями. Таким чином, можемо припустити, що дисперсії обох груп є статистично однаковими.

Гіпотеза 2. Про рівність середньої кількості фатальних аварій у робочі та вихідні дні.

Для перевірки гіпотези було обрано дні: четвер (що має “середні” значення відносно інших будніх днів і неділя).

Результати тестів Волда і Велча:

```
wald Test Statistic: 1.85921  
P-value (wald Test): 0.0314988>  
(welch Test):", p_value_welch)
```

```
welch Test Statistic: 1.85921  
P-value (welch Test): 0.0316464  
> |
```

Результати тестів і значення $p < 0.05$ в обох тестах вказують на статистично значущу різницю у середній кількості смертельних аварій між цими двома днями. Відхиливши нульову гіпотезу, ми можемо зробити висновок, що існує статистично значуща різниця між кількістю смертельних аварій у четвер та неділю.

Гіпотеза 3. Гіпотеза про рівність кількості фатальних аварій молодших та старших водіїв

Нульова гіпотеза (H_0): Середня кількість фатальних аварій на дорогах серед молодими та старшими водіями не відрізняється.

Альтернативна гіпотеза (H_1): Середня кількість фатальних аварій на дорогах серед молодими та старшими водіями відрізняється.

```
[1] "Chi-Squared Test Results:"  
[1] "Chi-Squared Statistic: 1861.80632455389"  
[1] "Degrees of Freedom: 12"  
[1] "p-value: 0"
```

Статистика Хі-квадрат: Значення статистики Хі-квадрат становить близько 1861.81. Це означає, що спостережувані частоти смертельних аварій для молодих і старших водіїв значно відрізняються від очікуваних частот, якби нульова гіпотеза (відсутність відмінностей у пропорціях смертельних аварій між віковими групами) була правильною.

Ступені свободи: Ми розрахували 12 ступенів свободи для цього тесту, враховуючи вікові категорії молодих і старших водіїв.

p-значення: Практично нульове p-значення вказує на те, ми маємо статистично значущі докази на користь відкидання нульової гіпотези.

Ці результати підтверджують, що молоді та старші водії мають значущі відмінності в пропорції смертельних аварій. Таким чином, можна зробити висновок, що молоді водії, ймовірно, демонструють вищу схильність до смертельних аварій, можливо, через їхню більш ризиковану поведінку та/або брак досвіду за кермом порівняно зі старшими водіями.

Гіпотеза 4. Про рівність фатальних аварій автомобілей з більшим та меншим віком за роками.

Для визначення зв'язку між віком автомобіля та кількістю фатальних аварій обчислимо коефіцієнт Спірмана

```
[1] "Коефіцієнт кореляції Спірмана між віком автомобіля і кількістю фатальними аваріями : -0.831011184486773"
```

Коефіцієнт кореляції Спірмана -0.831011184486773 вказує на те, що існує досить сильний зворотній взаємозв'язок між віком автомобіля та кількістю фатальних аварій. Це означає, що збільшення віку автомобіля супроводжується зниженням кількості фатальних аварій, і навпаки: молоді автомобілі більш схильні до фатальних аварій. Але це в свою чергу зумовлено більшою кількістю автомобілів молодшого віку на дорозі.

Якщо поділити вік авто на дві групи 1-10 та >10, то

```
"Коефіцієнт кореляції Спірмана між віком автомобіля і кількістю фатальними аваріями : -1"
```

Це означає дуже сильну негативну лінійну залежність між цими віковими групами та кількістю фатальних аварій.

Тест Волда

```
$mean_x  
[1] 18.52
```

```
$mean_y  
[1] 6.997143
```

```
$p_value  
[1] 1.61167e-05
```

```
$conf_int  
[1] 7.824452 15.221263
```

За результатами проведеного тесту перевірки гіпотези про рівність середніх значень двох груп даних (перші 10 років та другі 10), було виявлено статистично значущі відмінності між цими групами ($p < 0.05$). Середнє значення групи x складає 18.52, тоді як середнє значення групи y - 6.997143. Довірчий інтервал для різниці між середніми значеннями складає від 7.824452 до 15.221263. Таким чином, на підставі цих результатів можна зробити висновок про наявність статистично значущої різниці між двома

групами. На підставі отриманих результатів можна відкинути нульову гіпотезу про рівність середніх значень двох груп даних. Значення p-value ($1.61167e-05$) менше за зазначений рівень значимості (наприклад, 0.05), що свідчить про статистичну значущість різниці між групами. Таким чином, ми маємо достатні підстави вважати, що середні значення цих двох груп статистично відмінні одне від одного.

Тест Велча

```
Welch Two Sample t-test

data:  first_10_means and second_10_means

t = 6.755, df = 12.559, p-value = 1.612e-05

alternative hypothesis: true difference in means is not equal to 0

95 percent confidence interval:

 7.824452 15.221263

sample estimates:

mean of x mean of y
18.520000  6.997143
```

З результатів тесту Велча можна зробити наступні висновки: значення p-рівня статистичної значущості груп (перші 10 років та другі 10) дорівнює $1.612e-05$, що є дуже низьким значенням. Це свідчить про те, що різниця між середніми значеннями груп є статистично значущою. Довірчий інтервал для різниці між середніми значеннями груп складає (7.824452, 15.221263). Це означає, що ми можемо бути впевнені на рівні довіри 95%, що різниця між середніми значеннями груп лежить у цьому інтервалі. Значення t-статистики дорівнює 6.755. Це велике значення t-статистики підтверджує статистичну значущість різниці між середніми значеннями груп. Середнє значення групи перших 10-років дорівнює 18.520000, а середнє значення групи других 10-років дорівнює 6.997143. Це означає, що середнє значення групи перших 10-років суттєво вище, ніж середнє значення других 10-років. Отже, на підставі вищезазначених результатів ми можемо зробити висновок, що існує статистично значуща різниця між середніми значеннями груп перших 10-років і других 10-років.

Гіпотеза 5. Про рівність середньої кількості за роками фатальних аварій у міській та сільській місцевості .

Нульова гіпотеза (H_0): Середня кількість за роками фатальних аварій у сільській місцевості не відрізняється від середньої кількості фатальних аварій у міській місцевості.

Альтернативна гіпотеза (H_1): Середня кількість за роками фатальних аварій у сільській місцевості вища, ніж у міській місцевості.

Тест Волда:

```
$mean_x
[1] 192.4
$mean_y
[1] 65
$p_value
[1] 9.317599e-05
$conf_int
[1] 79.74555 175.05445
```

На підставі результатів тесту Волда, який показав дуже мале значення p -рівня ($9.317599e-05$), можна відхилити нульову гіпотезу про рівність середньої кількості фатальних аварій у сільській та міській місцевості. Значення p -рівня дозволяє прийняти альтернативну гіпотезу про те, що середня кількість фатальних аварій у сільській місцевості вища, ніж у міській. Довірчий інтервал для різниці між середніми також підтверджує цей висновок, показуючи статистичну значущість та великий розмір ефекту.

Тест Велча:

```
Welch Two Sample t-test

data: mean_fatal_rural and mean_fatal_urban
t = 5.8548, df = 11.471, p-value = 9.318e-05
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 79.74555 175.05445
sample estimates:
mean of x mean of y
 192.4      65.0
```

За результатами Welch Two Sample t-test можна зробити наступний висновок:

Отримане значення р-рівня ($9.318e-05$) є набагато меншим, ніж зазначений рівень значущості, що свідчить про статистичну значимість різниці у середніх кількостях фатальних аварій у сільській та міській місцевості. Отже, ми можемо відхилити нульову гіпотезу про рівність середніх.

Значення р-рівня, яке набагато менше встановленого рівня значущості, підтверджує альтернативну гіпотезу про те, що середня кількість фатальних аварій у сільській місцевості вища, ніж у міській.

Довірчий інтервал для різниці між середніми також підтверджує цей висновок, показуючи статистичну значущість та великий розмір ефекту, оскільки нижня межа довірчого інтервалу (79.74555) не перетинається з верхньою межею (175.05445).

Висновки

У ході цієї роботи ми провели статистичне виведення для різних гіпотез, що стосувалися факторів, які можуть впливати на серйозність і частоту дорожньо-транспортних пригод. Результати нашого аналізу дозволили нам встановити декілька ключових зв'язків:

1. Сезонність аварій: Виявлено статистично значущу різницю в кількості аварій між теплими та холодними місяцями.
2. Вплив дня тижня: Фатальні аварії частіше відбуваються у вихідні, що може бути пов'язано з більшою кількістю дорожнього руху та зміною характеру поїздок.
3. Вік водіїв: Молодші водії схильні до більшої кількості аварій, що свідчить про важливість освітніх програм для цієї категорії.
4. Вік транспортного засобу: Авто з меншим віком потрапляють частіше в ДТП.
5. Вплив місцевості: Виявлено, що фатальні аварії частіше відбуваються у сільській місцевості, ніж у міській.

Ці відкриття можуть бути використані для розробки цілеспрямованих заходів з підвищення дорожньої безпеки, включаючи інформаційні кампанії, поліпшення інфраструктури та зміну дорожнього законодавства, щоб зменшити частоту та серйозність аварій.

Список літератури

1. Д. Ю. Тавров. «Аналіз даних» - електронний конспект. Київ.