

Anàlisi de la perillositat dels accidents de tràfic a Catalunya

David Candela Rubio (1563873)

16 de Juny de 2021

Índex

Resum	1
1 Introducció	1
2 Dades	1
3 Moment	1
3.1 Hora del dia	1
3.2 Dia de la setmana	2
3.3 Època de l'any	3
4 Lloc	4
4.1 Tipus de via	4
4.2 Tipus de carril	5
5 Altres	9
5.1 Velocitat màxima de la via	9
5.2 Visibilitat	11
5.3 Condicions del terreny	12
6 Conclusions	12
7 Annex	13
7.1 Dades	13
7.2 Altres Gràfiques	14
8 Apèndix	16
8.1 Codi d'R	16
Referències	16

Anàlisi de la perillositat dels accidents de tràfic a Catalunya

Resum

Estudiarem quins són els factors principals en els accidents de tràfic més perillosos. Primer estudiarem els factors corresponents al temps, en concret la franja horària amb major risc, la diferència entre els dies de la setmana, i si hi ha alguna època de l'any en concret on la perillositat augmenti o està dispers per tot l'any. Després sobre el lloc mirarem la via, si alguna té risc major o menor, i quins carrils que tenen alguna funció especial (bici, carril lent, carril d'avançament ...) no són tant o ho són més de perillosos que els carrils normals. Finalment mirarem altres factors com la correlació entre la velocitat i la perillositat, el canvi del risc amb diversos aspectes que poden afectar la visibilitat, i l'estat del terreny (sec, mullat, nevat ...) en el moment de l'accident.

1 Introducció

Cada dia hi ha accidents de tràfic, a vegades ningú en resulta ferit en d'altres alguns dels implicats en surten ferits i, desgraciadament, a vegades alguns dels implicats acaben morint en l'accident o poc després d'aquest.

En aquest estudi analitzarem quan, on i quins factors estan relacionats amb la perillositat en els accidents de tràfic, és a dir, si hi hagués un accident de tràfic quantes persones acabarien ferides i quantes mortes. Per mesurar la perillositat dels accidents es poden fer servir diverses mètriques, les principals essent el número de víctimes que va haver en l'accident i el número de morts que va haver en l'accident.

Utilitzant aquestes dues mètriques analitzaré quins factor es troben relacionats amb la perillositat dels accidents. Per fer-ho utilitzaré mètodes com F-test, test permutacional, Bootstrap i Bootstrap no paramètric

2 Dades

Per fer l'anàlisi de la perillositat dels accidents he agafat de la pàgina de la Generalitat de Catalunya [dades obertes](#) [1], les dades corresponents als accidents de tràfic entre 2010 i 2018 ([Accidents de trànsit amb morts o ferits greus a Catalunya](#) [2]). Com que el conjunt de dades és només d'accidents on hi ha ferits o morts la mesura de perillositat de víctimes per accident serà sempre com a mínim 1. Es pot trobar totes les dades junt amb el codi en R (R markdown) al repositori [UNI_AddC.Exam](#) [3].

El conjunt de dades conté informació sobre la data, l'hora, el lloc (carretera, municipi, carril, punt kilomètric ...), les condicions presents durant l'accident (climatologia, visibilitat ...), raons que han influït a l'accident, número de persones implicades (ferits, greus, morts), vehicles implicats (tipus i número) ... Abans de ser utilitzat hem fet un seguit d'alteracions, i correccions per poder utilitzar les dades amb més facilitat. Aquestes modificacions i la descripció de les Dades utilitzades per l'anàlisi es troba a l'Annex, l'explicació de totes les dades del conjunt de dades original es pot trobar a la pàgina d' [Accidents de trànsit amb morts o ferits greus a Catalunya](#) [2])

3 Moment

De les diferents dades que corresponen al temps només analitzaré les referents a l'hora del dia, el dia de la setmana i el mes. He escollit aquestes perquè són les que podrien tenir un efecte en el número de víctimes i morts per accident.

Altres dades com el número de dia del mes no hi hauria cap raó aparent per que hi influís, els minuts no afectarien massa a l'hora i la seva distribució mostra que tan sols són aproximacions amb diversos nivells de precisió (Distribució dels minuts).

3.1 Hora del dia

La meua hipòtesis és que els accidents seran més perillosos entre les 23:00 i les 6:00, ja que la gent estaria més cansada i la visibilitat seria menor, per tant, el temps i la velocitat per poder reaccionar i evitar o reduir la

gravetat d'un accident seria menor, fent els accidents produïts a aquesta franja amb una mortalitat i un número de persones afectades major.

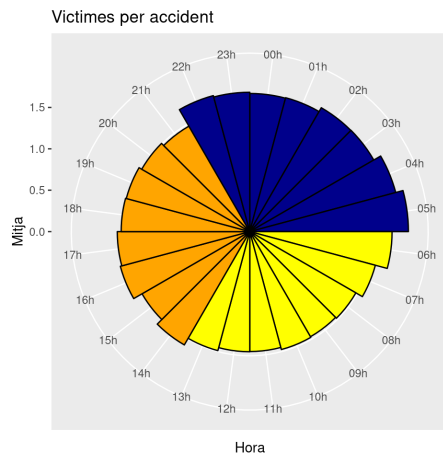


Figura 1: Víctimes
Mitja de víctimes a cada hora
Blau per la nit, groc pel matí i taronja per la tarda

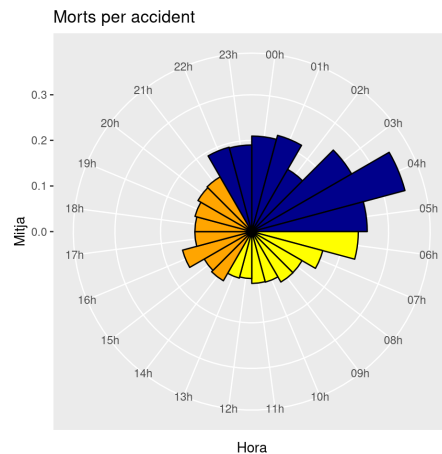


Figura 2: Morts
Mitja de morts a cada hora

Després de veure les gràfiques sembla que hi ha una pujada entre les 22:00 i les 6:00. Per comprovar si aquesta diferència és significativa faig un test permutacional que consisteix en fer diverses permutacions de les dades originals i analitzar els aspectes que ens importen en les diferents permutacions generades. Aquest test el fem per calcular la mitja i la variància circular i d'aquesta manera comprovar si aquest canvi és realment notable. (Comprovem que la variància sigui considerablement gran, que seria equivalent a que els valors d'una franja són significativament major)

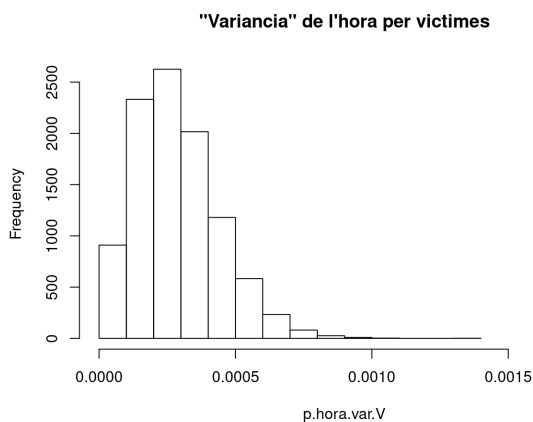


Figura 3: Variància Víctimes
Variància circular per víctimes

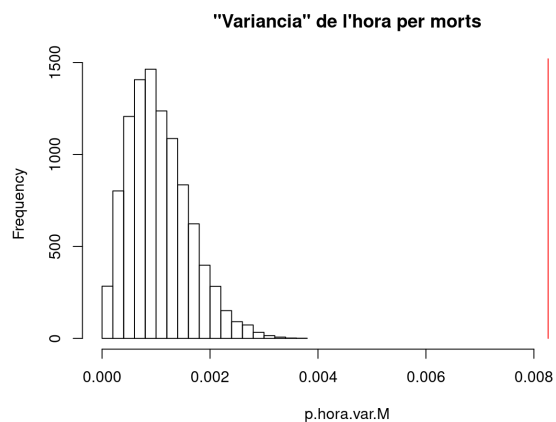


Figura 4: Variància Morts
Variància circular per morts

	p-value
Víctimes	0
Morts	0

Després de fer el test permutacional amb un total de 10000 repeticions podem comprovar que no s'arriba a tenir una variància tan gran en cap de les permutacions, llavors es pot afirmar que realment, en ambdós casos, la gravetat dels accidents que es produeixen puja al voltant de les $2:00 \pm 4h$.

3.2 Dia de la setmana

Analitzem també els dies de la setmana perquè el fet de ser festiu o laborable pot afectar a la quantitat de gent que condueix, la quantitat de gent que va en cada cotxe i el comportament d'aquesta (diferents destinacions de les normals ...). Llavors seria possible veure un canvi a dissabte i sobretot diumenge (ja que diumenge és festiu per més persones)

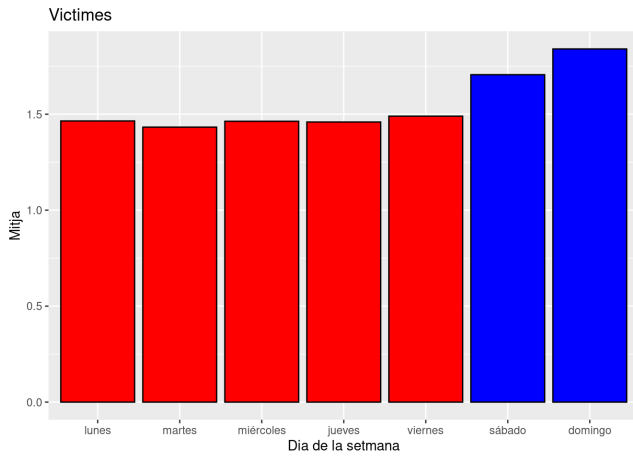


Figura 5: Víctimes

Mitja de víctimes als diferents dies de la setmana

Blau pel cap de setmana, vermell pels dies laborables

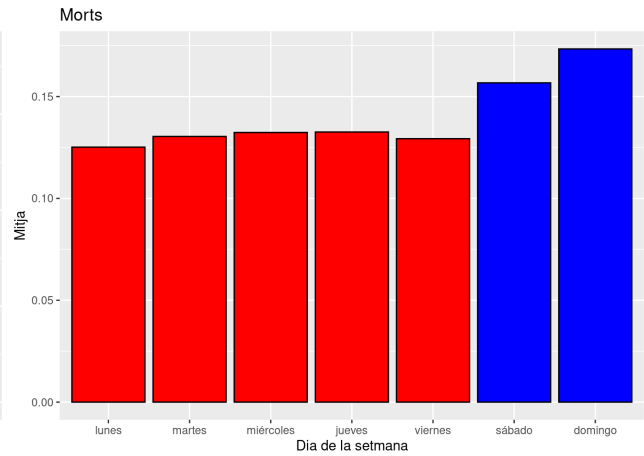


Figura 6: Morts

Mitja de morts als diferents dies de la setmana

Amb les gràfiques del número víctimes i morts per accident veiem que, en efecte, sembla haver-hi aquest canvi de perillositat el cap de setmana, fent-se més notable el diumenge. Per assegurar-nos de que aquests tres canvis són significatius he fet tres F-test (test ANOVA per comprovar diverses hipòtesis a la vegada amb la instrucció aov).

El primer:

\mathcal{H}_0 : Els dies laborables no són iguals de perillosos

\mathcal{H}_1 : Els dies laborables tenen la mateixa perillositat

El segon:

\mathcal{H}_0 : Dissabte és més perillós que els dies laborable

\Rightarrow Diumenge és més perillós que els dies laborable

\mathcal{H}_1 : Dissabte és tan perillós com els dies laborable

El tercer:

\mathcal{H}_0 : Diumenge és més perillós que Dissabte

\mathcal{H}_1 : Diumenge és tan perillós com Dissabte

	p-value	
	Víctimes	Morts
$test_1$	0.4398894	0.961798
$test_1$	2.363101e-23	0.00154988
$test_3$	0.002842418	0.2385609

Amb això podem descartar que els dies laborables tinguin una perillositat diferent i acceptar que el cap de setmana té una major perillositat que els dies laborables. Finalment, depenent de si estem mirant víctimes o morts acceptem o rebutjem la tercera hipòtesis respectivament.

3.3 Època de l'any

Per les mateixes raons que pels dies de la setmana és possible que hi hagi un increment de la perillositat cap a l'estiu ja que molta gent té vacances durant aquesta estació, tot i això és possible que no acabi sent significatiu degut a que també hi ha festes per nadal, cap d'any, setmana santa ...

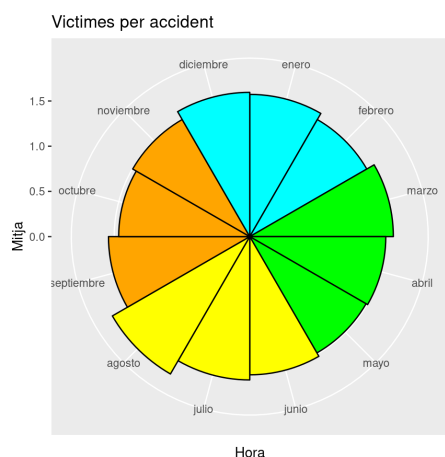


Figura 7: Víctimes
Mitja de víctimes durant l'any
Blau per l'hivern, verd per la primavera, groc per l'estiu i taronja per la tardor

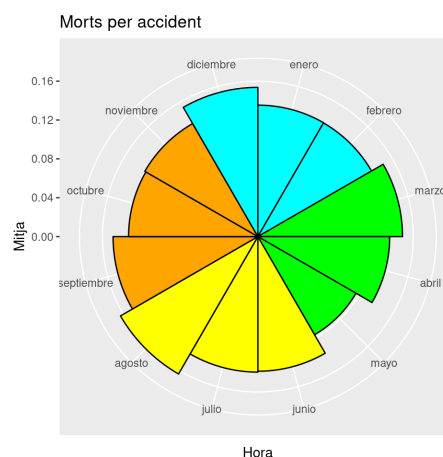


Figura 8: Morts
Mitja de morts durant l'any
Blau per l'hivern, verd per la primavera, groc per l'estiu i taronja per la tardor

Al veure les gràfiques es veu l'increment per l'estiu (a finals d'estiu inicis de tardor) però també es nota les festes nadalenques i de setmana santa (sobre tot a les morts). Per comprovar si arriba a haver algun increment cap alguna època de l'any en concret, ho analitzaré igual que amb l'hora, amb un test permutacional calculant la mitja i la variància circular amb un total de 10000 permutacions.

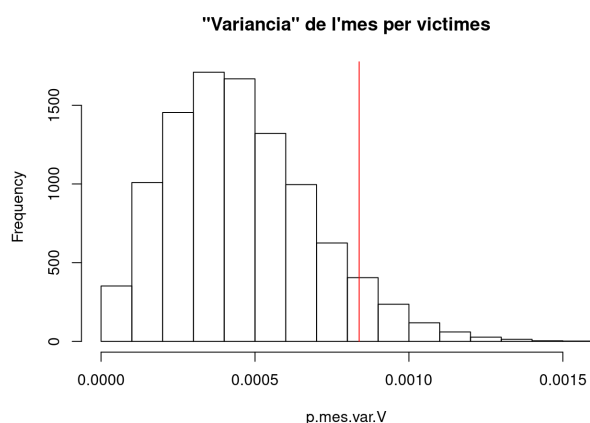


Figura 9: Variància Víctimes
Variància circular per víctimes

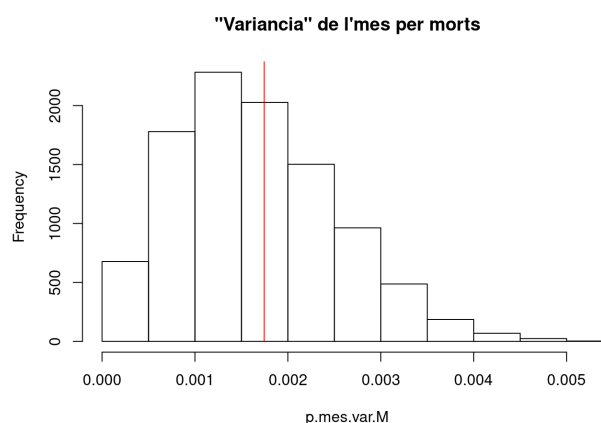


Figura 10: Variància Morts
Variància circular per morts

	p-value
Víctimes	0.0622
Morts	0.4254

Tot i que per víctimes el p-value és més petit que per morts, no ho és suficient per tal de que els resultat siguin significatius. Per tant no hi ha una única època de l'any en concret on la perillositat augmenti sinó que sembla estar repartit entre les diverses festivitats al llarg de l'any.

4 Lloc

Pel lloc analitzarem el tipus de via i el carril en concret on es produeix l'accident (en cas de que aquest tingui alguna funcionalitat especial). En concret mirarem si per algun d'aquest la perillositat serà major o menor que la resta.

4.1 Tipus de via

Com que les vies urbanes i els camins rurals forcen a baixar la velocitat seria d'esperar que ambdós fossin els menys perillosos de tots.

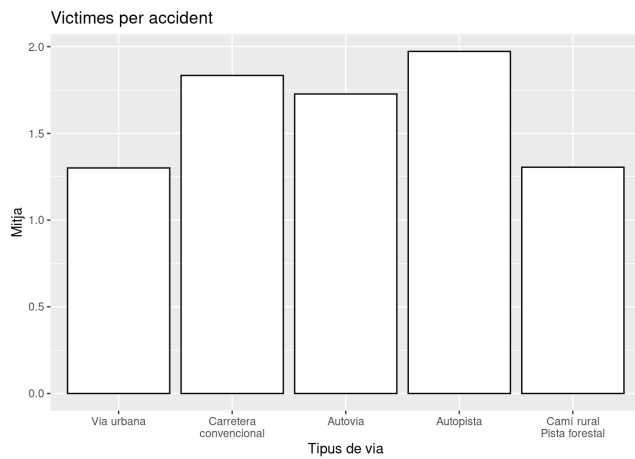


Figura 11: Víctimes
Mitja de víctimes als diferents tipus de via

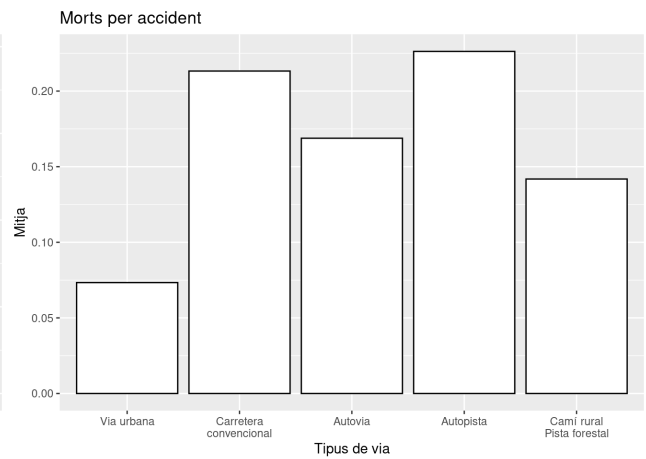


Figura 12: Morts
Mitja de morts als diferents tipus de via

Havent fet un gràfic de les dades sembla que si que ho son per víctimes però només les vies urbanes per morts. Per comprovar-ho he calculat uns intervals de confiança amb el mètode de Bootstrap no parametric el qual consisteix a fer diverses mostres amb reemplaçament de l'original i analitzar les dades que es volen mesurar en aquestes noves mostres.

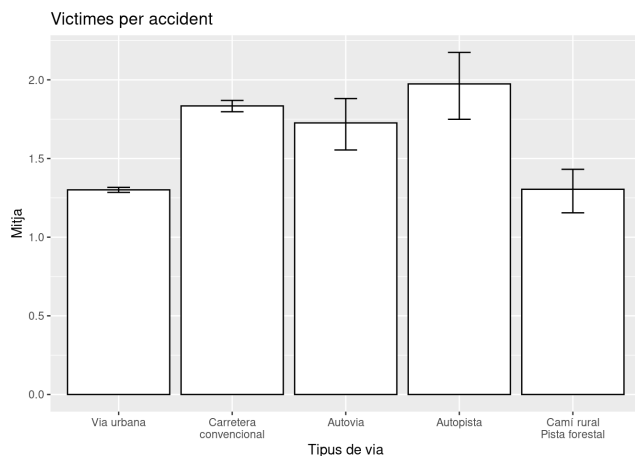


Figura 13: Víctimes
Mitja de víctimes als diferents tipus de via amb interval de confiança

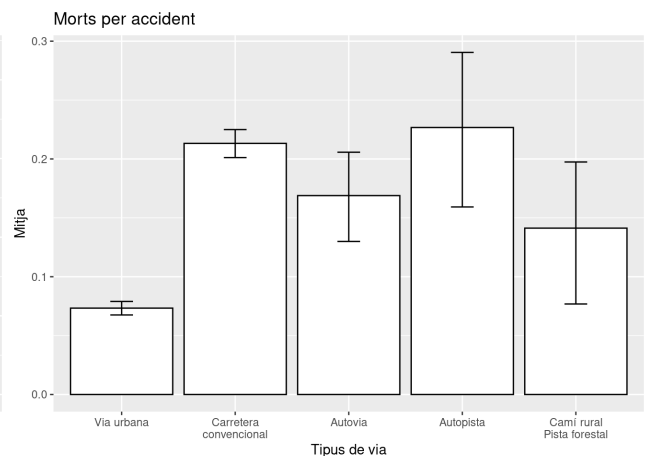


Figura 14: Morts
Mitja de morts als diferents tipus de via amb interval de confiança

Tal i com pensàvem, per víctimes via urbana i camí rural tenen un número menor de víctimes per accident. Pel que fa a morts per accident, podem confirmar que a les vies urbanes aquest valor és menor però pels camins rurals el seu interval de confiança és massa gran per afirmar o negar que és menys mortal que els altres. Finalment també es pot observar que la resta de vies coincideixen en els seus intervals de confiança.

4.2 Tipus de carril

La hipòtesis inicial és que el carril bici serà el menys perillós perquè ara acostuma a separar-se una mica dels carrils dels cotxes donant més temps al conductor per evitar o reduir els danys de l'accident a més de que les úniques víctimes que probablement hi hagi serà el ciclista. El més perillós probablement serà el d'avançament ja que en aquest carril s'accelerà augmentat els danys dels accidents que es puguin produir en aquest.

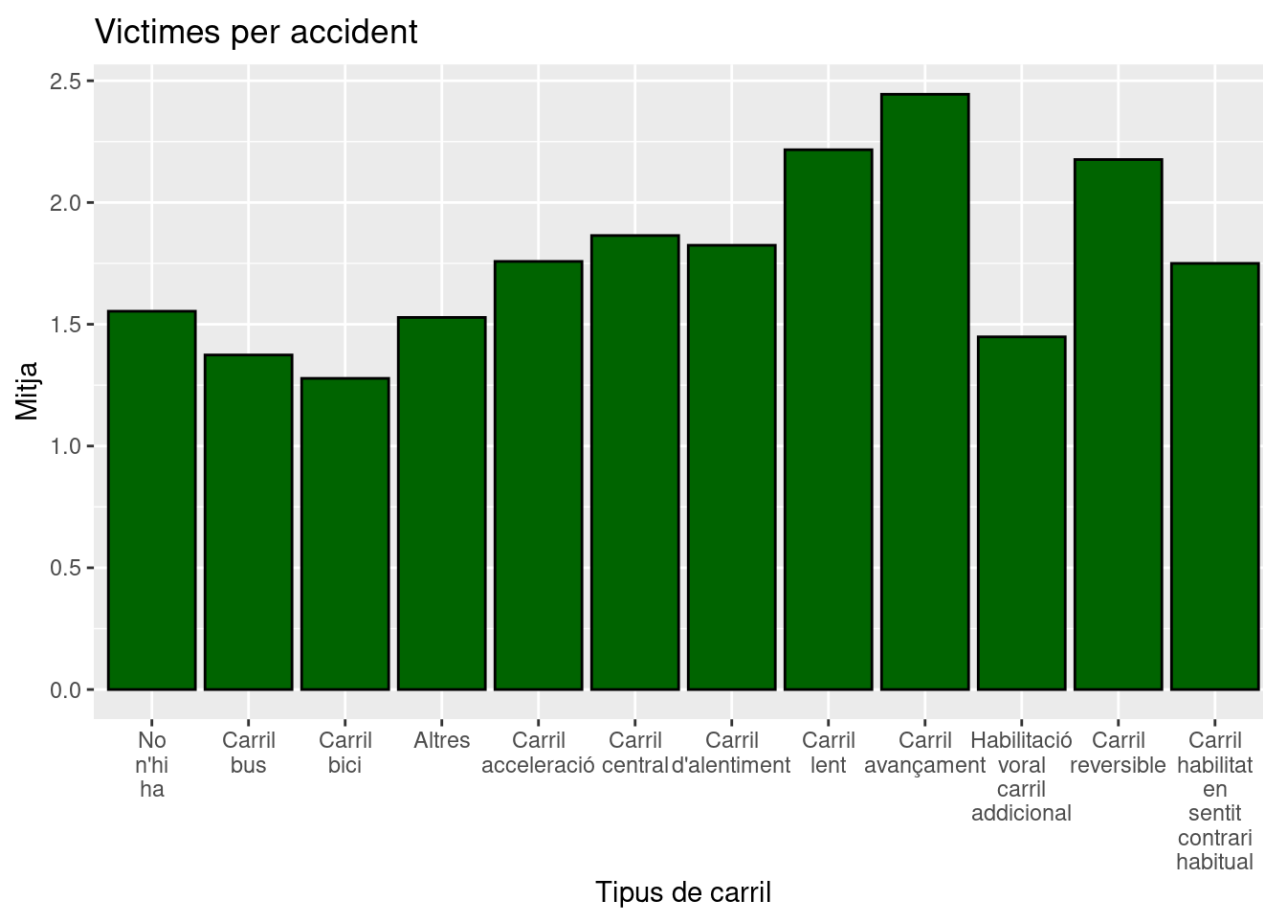


Figura 15: Víctimes
Mitja de víctimes als diferents tipus de carril

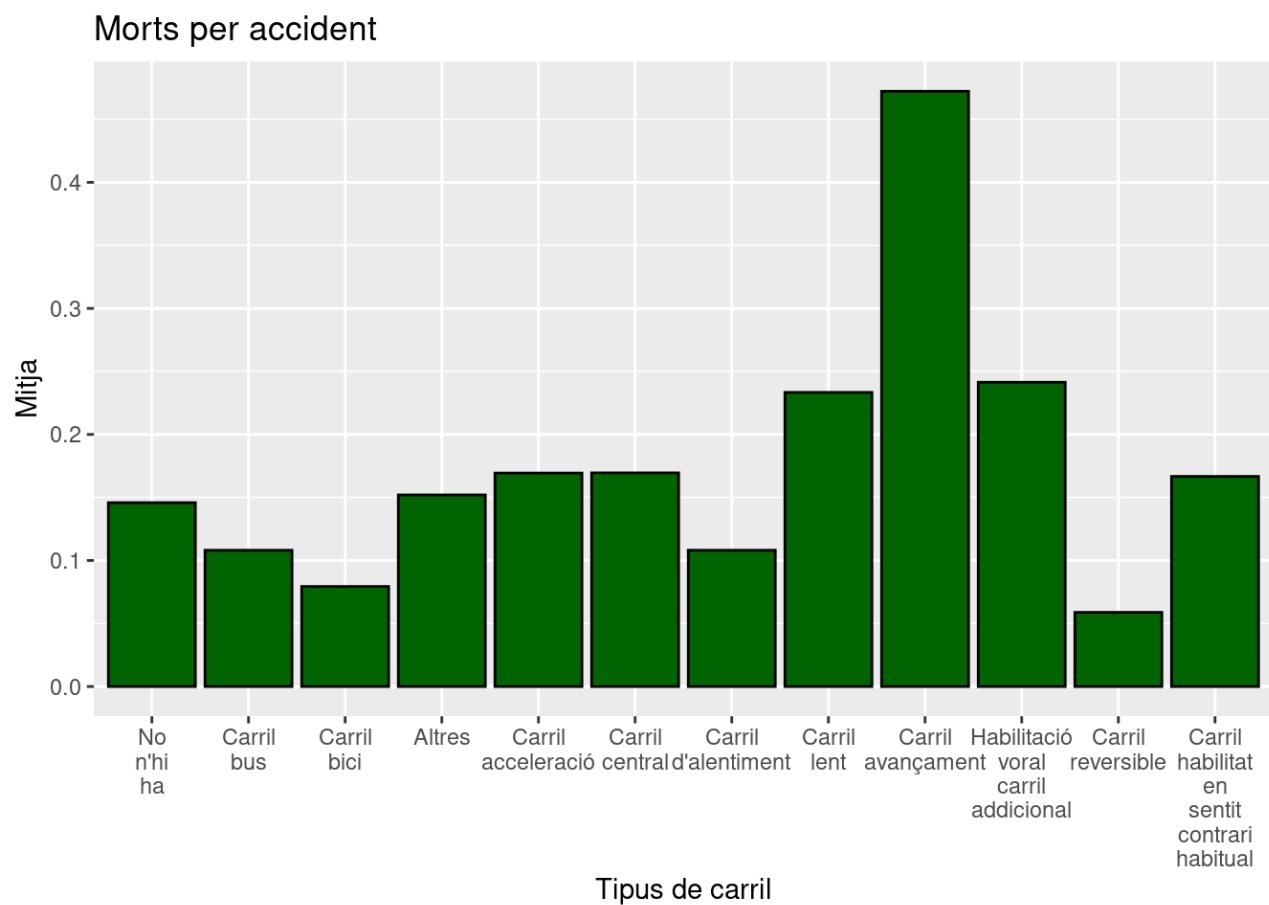


Figura 16: Morts
Mitja de morts als diferents tipus de carril

Per analitzar les dades he decidit fer intervals de confiança per cada tipus de carril i després comprovar-ne quins superen o es troben per sota del cas on els carril no té funció especial. Per calcular els intervals de confiança he utilitzat el mateix mètode que a Tipus de via.

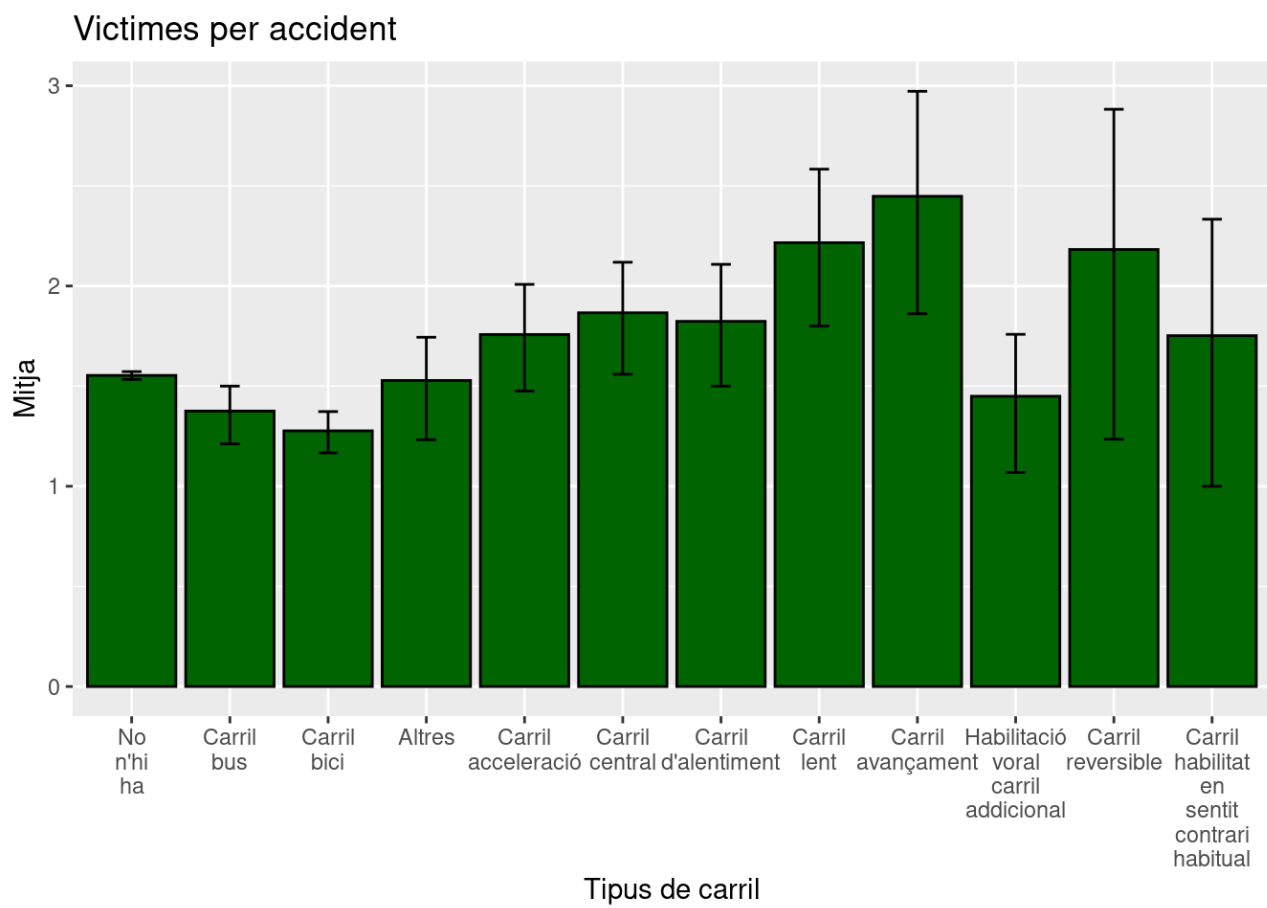


Figura 17: Víctimes
Mitja de víctimes als diferents tipus de carril amb interval de confiança

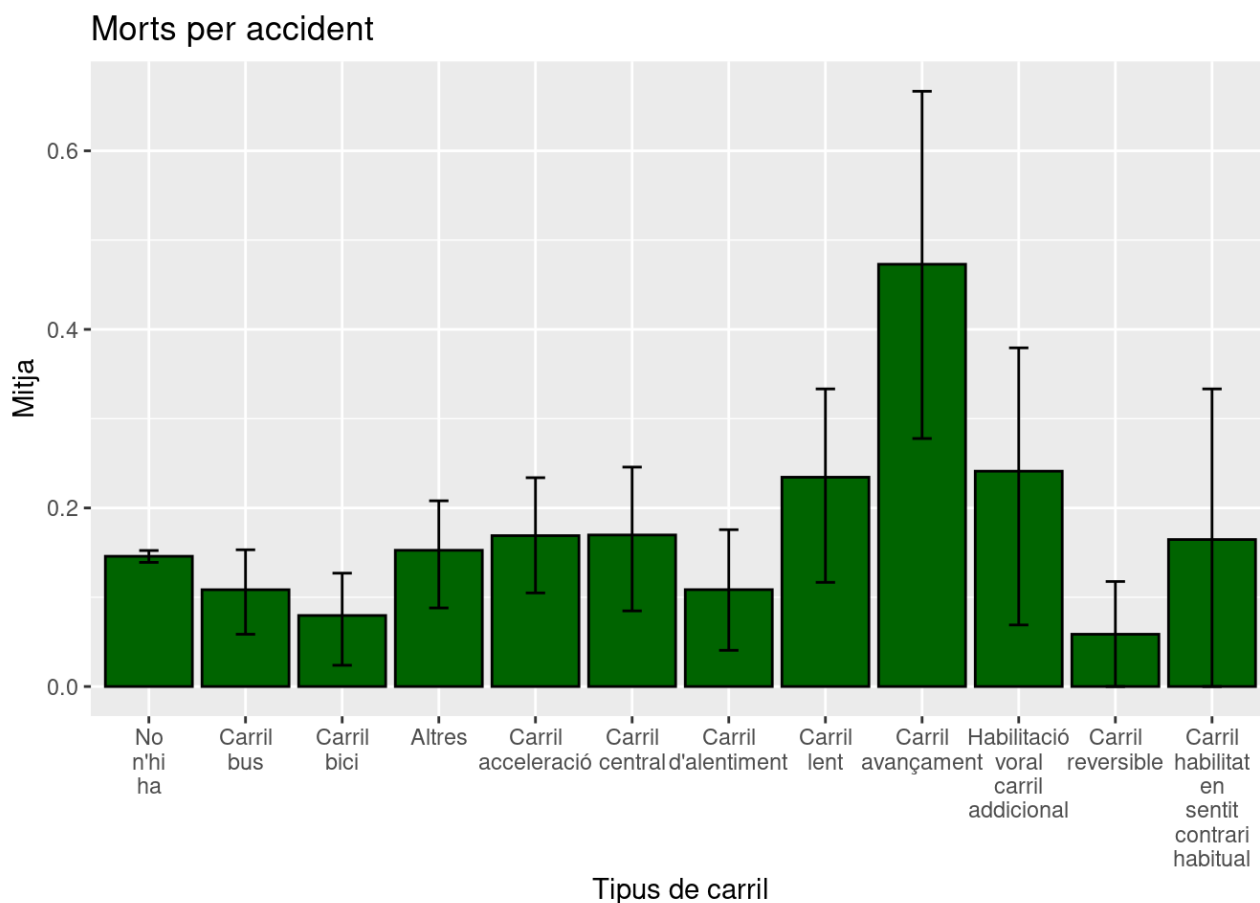


Figura 18: Morts
Mitja de morts als diferents tipus de carril amb interval de confiança

Pel que fa a víctimes els únics carrils que no coincideixen en l'interval de confiança amb el cas on el carril no té cap funció especial són el carril lent i el d'avançament que el superen i el de bus i bici que es troben per sota. Pels morts l'únic que el supera és el d'avançament i els dos que es troben per sota són el de bici i el reversible. Tota la resta de tipus de carrils coincideixen en l'interval amb l'opció de carril sense funció especial, per tant les seves diferències no són significatives.

Llavors el carril més perillós és el d'avançament i el menys perillós el bici.

5 Altres

5.1 Velocitat màxima de la via

Quan major és la velocitat menys temps pot haver-hi per reaccionar i l'impacte de l'accident serà major, per tant s'esperaria que la velocitat estigués correlacionat positivament amb el número de víctimes als accidents, el número de morts als accidents i el número de víctimes que moren en els accidents.

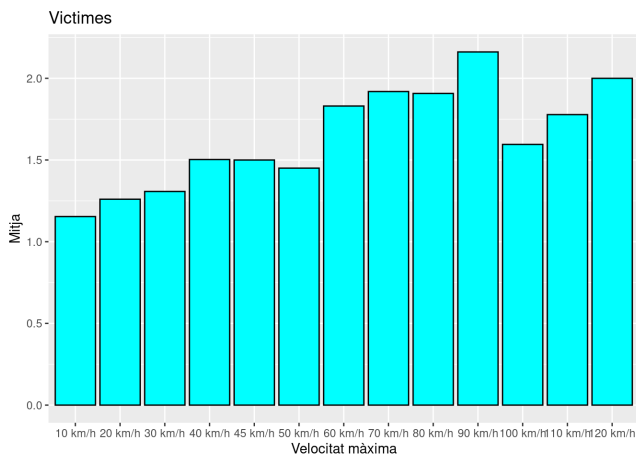


Figura 19: Víctimes
Mitja de víctimes per les diferents velocitats màximes permeses

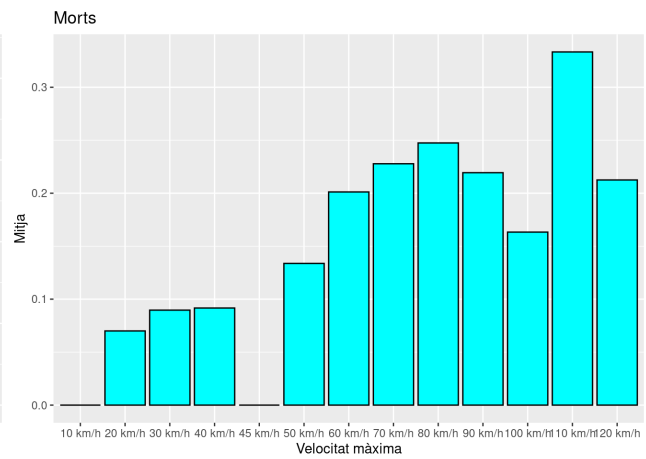


Figura 20: Morts
Mitja de morts per les diferents velocitats màximes permeses

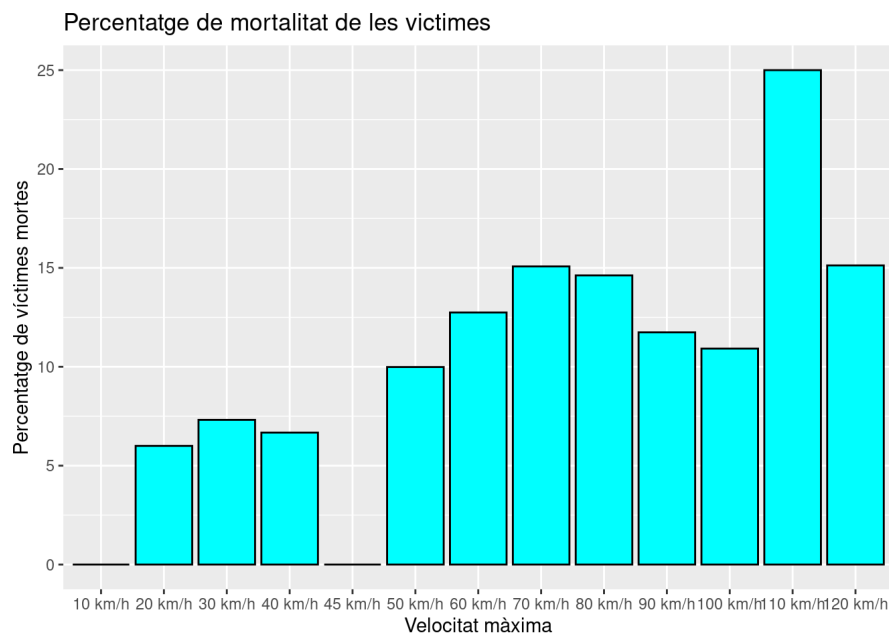


Figura 21: Morts per víctima
Percentatge de víctimes mortes per les diferents velocitats màximes permeses

Faré un test permutacional calculant els coeficients de pearson (correlació lineal) i de spearman (correlació monòtona) per comprovar si la correlació entre les dades és positiva.

Els resultats d'aquest són els següents:

	Víctimes		Morts		Mortalitat	
	correlació	p-value	correlació	p-value	correlació	p-value
Pearson	0.04008627	0	0.04403149	0	0.0351944	0
Spearman	0.01377029	0.057	0.0318193	0.001	0.03048427	0.001

En tots els casos ens surt que estan correlacionats linealment positivament amb la velocitat (valor de correlació de pearson major a 0 i p-value menor a 0.025), però només el número de morts per accident i el número de víctimes mortes tenen una relació monòtona positiva. (valor de correlació de spearman major a 0 i p-value menor a 0.025)

5.2 Visibilitat

Respecte a la boira, esperàriem que els accidents no fossin tan perillosos, com que la boira no és un fenomen freqüent, la gent probablement anirà amb major precaució a l'hora de conduir, reduint així la perillositat, per tant esperàriem veure un decrement quan hi ha boira (en cas de haver-hi una diferència). En quant a si és de dia o de nit i a la il·luminació, a diferència de la boira i com s'ha vist abans en l'anàlisi de Hora del dia, per la nit serà major i esperàriem que amb pitjor il·luminació també augmentes la perillositat.

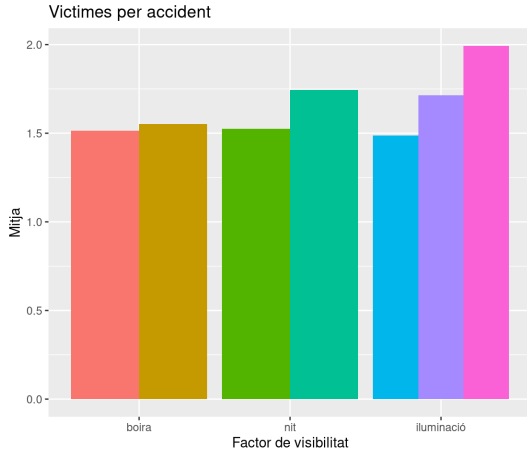


Figura 22: Víctimes

Mitja de víctimes per les diferents condicions de visibilitat

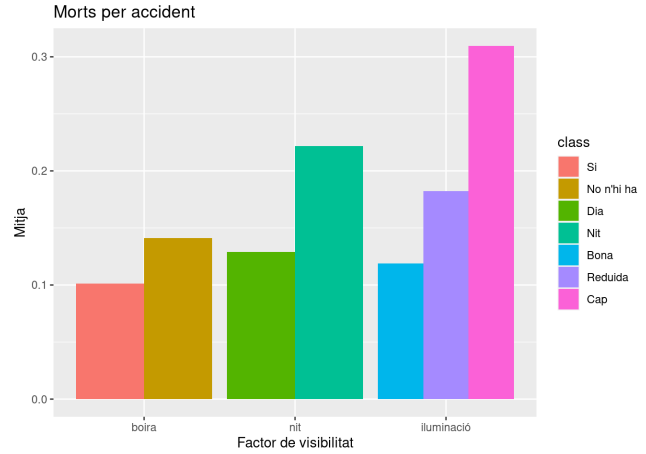


Figura 23: Morts

Mitja de morts per les diferents condicions de visibilitat

A primera vista sembla que els resultats són els que s'havia previst amb la hipòtesis inicial, però per assegurar-nos encara hem de comprovar si les diferències són significatives o no. Si calculem els intervals de confiança podem comprovar si les hipòtesis són certes o no mirant si els diversos intervals se solapen entre ells (és possible que aquest factor no afecti) o no (la hipòtesis inicial és certa, el factor i la perillositat estan relacionats). Per calcular els intervals de confiança he utilitzat el mètode de Bootstrap que consisteix en generar diverses mostres amb les característiques de l'original i analitzar les dades que ens interessen en aquestes noves mostres.

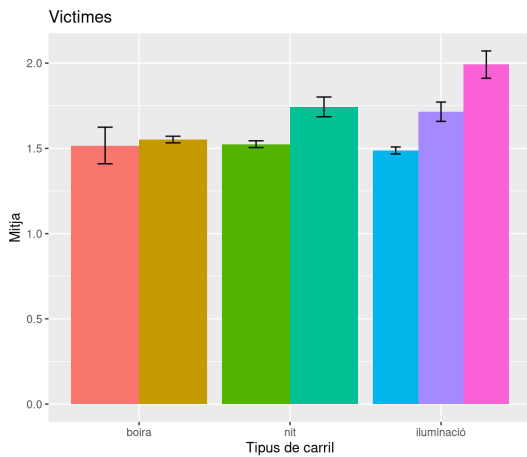


Figura 24: Víctimes

Mitja de víctimes per les diferents condicions de visibilitat amb intervals de confiança

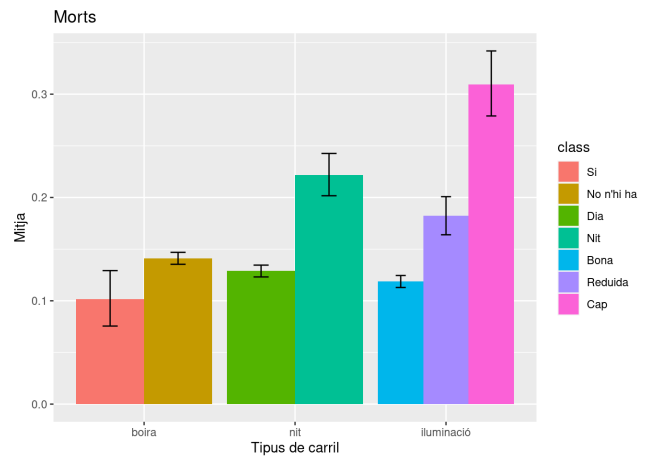


Figura 25: Morts

Mitja de morts per les diferents condicions de visibilitat amb intervals de confiança

Amb els valors dels intervals de confiança:

Factor	Boira		Dia/Nit		Il·luminació		
	Si	No n'hi ha	Dia	Nit	Bona	Reduïda	Cap
Mínim (víctimes)	1.409543	1.532481	1.504164	1.686458	1.466888	1.659554	1.913265
Màxim (víctimes)	1.624254	1.571262	1.544377	1.800799	1.507912	1.771581	2.07568
Mínim (morts)	0.075547	0.135333	0.123011	0.201198	0.113008	0.163919	0.278912
Màxim (morts)	0.131213	0.146949	0.134453	0.242176	0.124538	0.200776	0.340986

Per tant podem veure que les hipòtesis per si és de dia o nit i de la il·luminació són certes, llavors aquests dos fets es troben relacionats amb el número de víctimes i el de morts. Però en quant a la boira, només la diferència respecte el número de morts és significativa.

5.3 Condicions del terreny

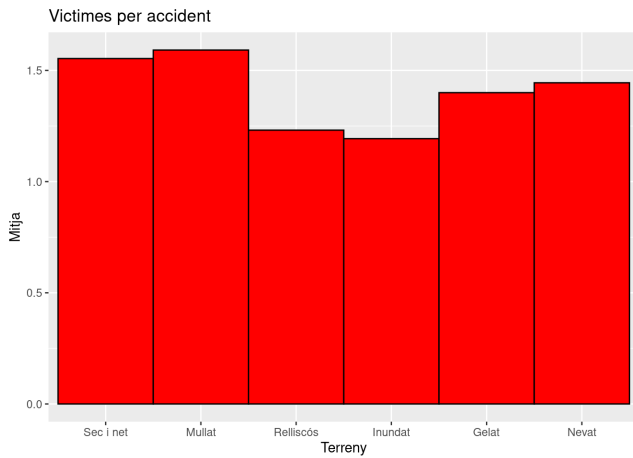


Figura 26: Víctimes

Mitja de víctimes per les diferents condicions del terreny

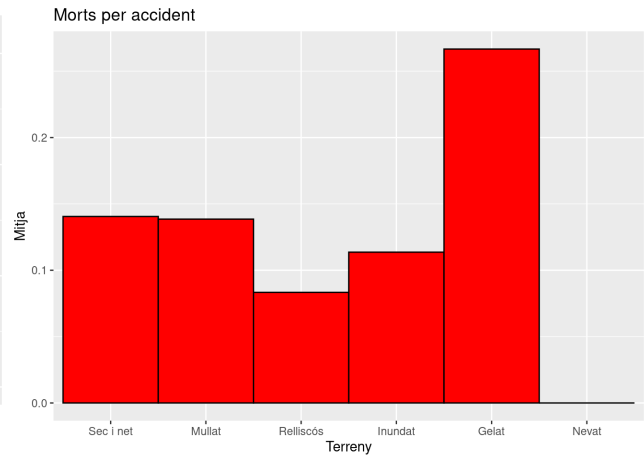


Figura 27: Morts

Mitja de morts per les diferents condicions del terreny

Per comprovar si algun (en concret terreny relliscós i inundat) és diferent a la resta calcularé els intervals de confiança de si tots fossin iguals (\mathcal{H}_0) amb Bootstrap no parametric de manera similar a com ho he fet a Tipus de via i Tipus de carril. D'aquesta manera, si algun es troba fora podrem comprovar que la seva perillositat és diferent. (\mathcal{H}_1)

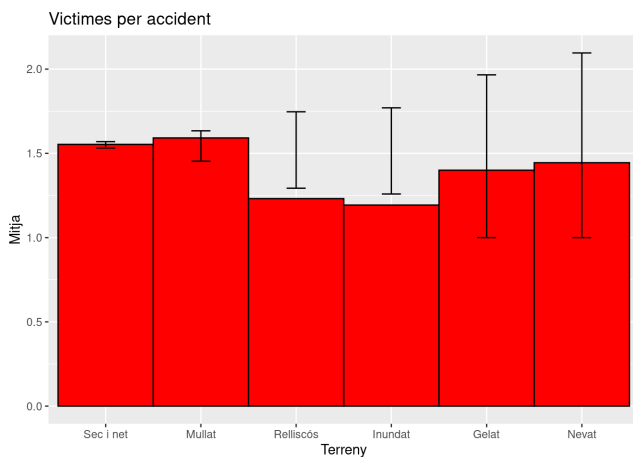


Figura 28: Víctimes

Mitja de víctimes per les diferents condicions del terreny i intervals de la hipòtesis nula

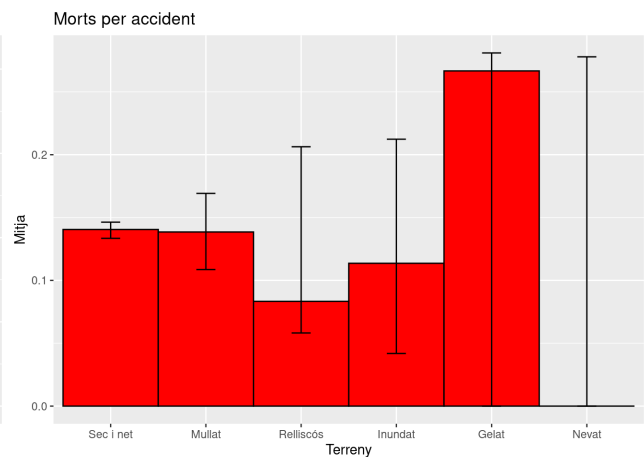


Figura 29: Morts

Mitja de morts per les diferents condicions del terreny i intervals de la hipòtesis nula

Amb aquests resultats es pot concloure que, pel que fa a víctimes per accident, només en els casos de terreny relliscós i inundat aquest valor és pot afirmar que és diferent, en concret és menor que la resta. Pel que fa als morts per accident cap resultat es troba massa allunya per tal de ser significatiu.

6 Conclusions

Els factors comuns dels accidents més perillosos són una il·luminació dolenta, duarnt la nit (23:00-6:00), en els cap de setmana, en sec net, alta velocitat.

També podem extreure que en situacions poc comuns, encara que pugin semblar més perilloses, la gent anirà amb major precaució i els danys produïts en un accident probablement seran iguals o menors. En canvi si són situacions que passen freqüentment la gent no anirà amb suficient precaució i els resultats d'un accident llavors serà major als altres casos.

7 Annex

7.1 Dades

Llista de dades que s'utilitzen per l'anàlisi junt amb una petita explicació

1. Mesurar l'índex de perillositat
 - (a) F_VICTIMES: número total de ferits i morts a l'accident
 - (b) F_MORTS: número de morts en l'accident
2. Hora del dia

Com el format de l'hora no era consistent (p.e. 0:15 guardat com 15, 1:50 guardat com 15, 15:00 guardat com 15 ...) s'ha fet el millor possible per intentar recuperar l'hora original de l'accident

 - (a) hor: hora de l'accident
corregit respecte l'hora original
 - (b) min: minut aproximat de l'accident
extret a partir del valor de l'hora original
 - (c) grupHor: marca si l'accident s'ha produït pel matí, la tarda o la nit
3. Data
 - (a) data: data de l'accident en format "dd/mm/yyyy"
 - (b) diaSem: nom del dia de la setmana de l'accident
extret a partir del valor de la data
 - (c) mes: nom del mes de l'accident (extra)
extret a partir del valor de la data
4. Via
 - (a) D_TIPUS_VIA: marca si l'accident s'ha produït a una autopista, autovia, carretera convencional ...
 - (b) D_CARRIL_ESPECIAL: marca si el carril de l'accident tenia alguna funció especial o no, i en cas de tenir-ne l'especifica (p.e. avançament, bici ...)
 - (c) C_VELOCITAT_VIA: Velocitat màxima permesa a la via de l'accident
S'han canviat els valors on s'havia omplert amb un valor incorrecte (p.e. 999, 0 ...) per NA per tal de poder treballar amb més facilitat
5. Estat de visibilitat
 - (a) D_BOIRA: marca si en el moment de l'accident hi havia boira
 - (b) D_LLUMINOSITAT: diu com era la il·luminació en el moment de l'accident
6. Estat de la carretera
 - (a) D_SUPERFICIE: condicions de la carretera en el moment de l'accident (p.e. mullat, reliscós, sec)

7.2 Altres Gràfiques

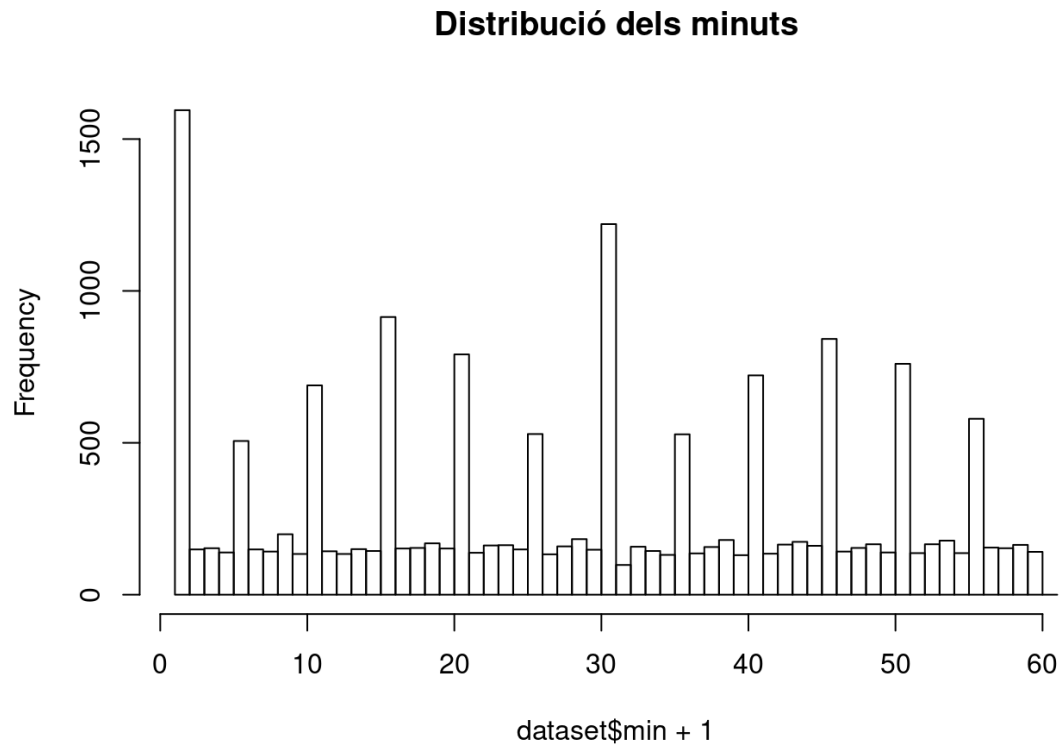


Figura 30: Distribució dels minuts

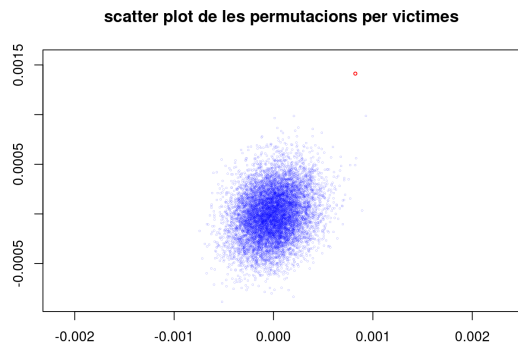


Figura 31: Víctimes
Distribució de la mitja circular de les
permutacions per víctimes

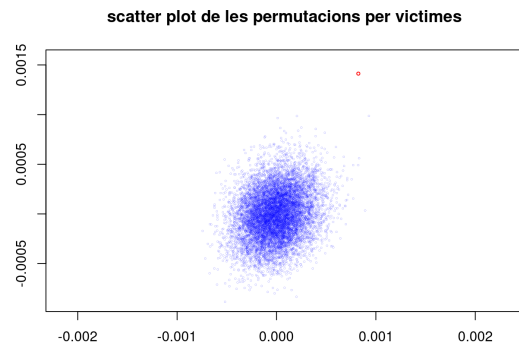


Figura 32: Morts
Distribució de la mitja circular de les
permutacions per morts

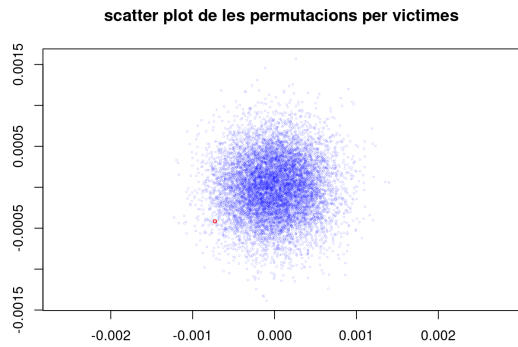


Figura 33: Víctimes
Distribució de la mitja circular de les
permutacions per víctimes

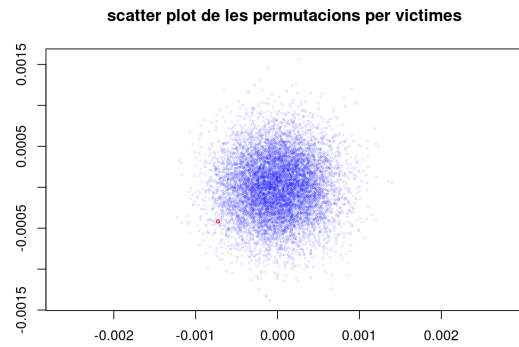


Figura 34: Morts
Distribució de la mitja circular de les
permutacions per morts

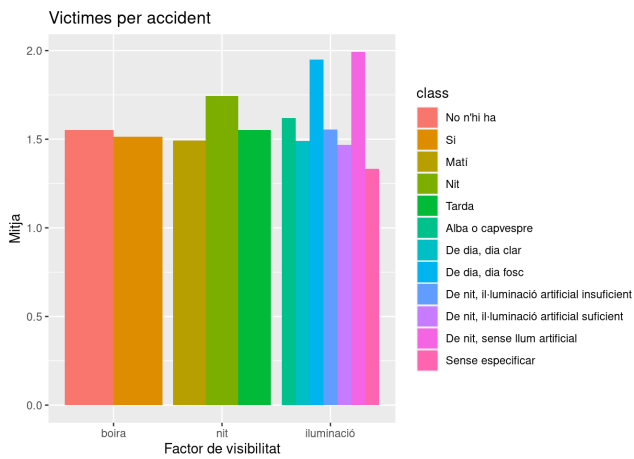


Figura 35: Víctimes
Víctimes per accident amb totes les característiques
originals

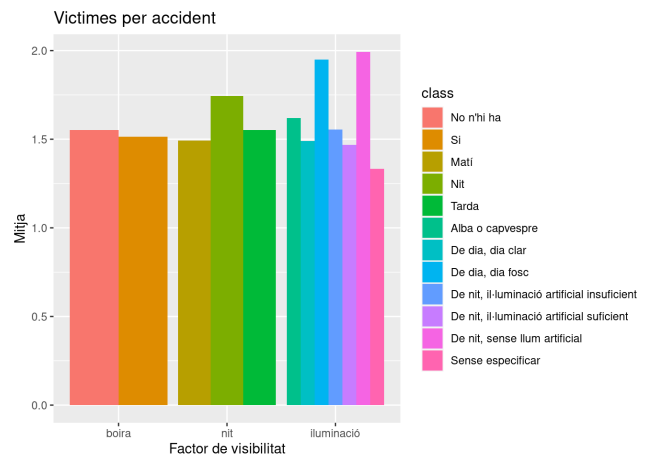


Figura 36: Morts
Morts per accident amb totes les característiques
originals

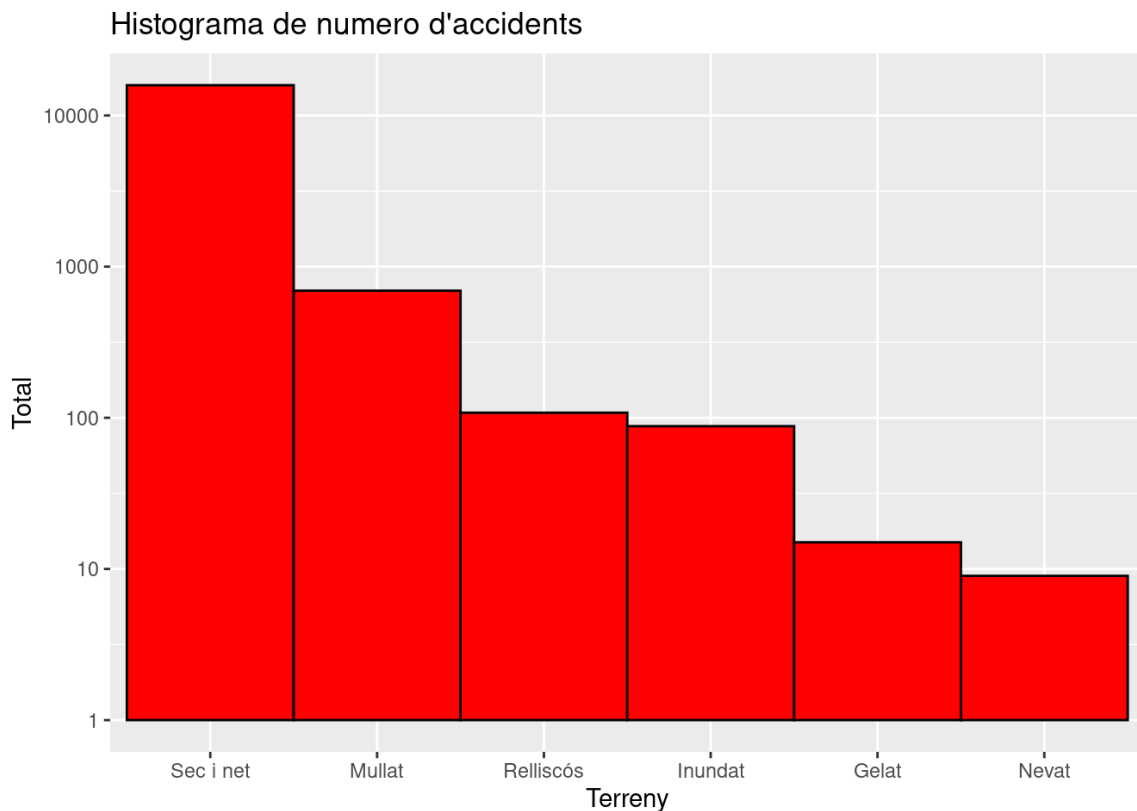


Figura 37: número d'accidents depenent del tipus de terreny (escala log-y)

8 Apèndix

8.1 Codi d'R

Com que el codi d'R té més de 900 línies de codi el deixaré adjunt en un altre fitxer anomenat 'R.Code.Rmd' junt amb el fitxer 'R.Code.html' on també es poden veure els resultats de l'execució d'aquest. El codi d'R també és pot accedir a través del [repositori de GitHub](#) [3]

[Enllaç](#) al fitxer del codi d'R en R markdown. (dins del repositori de GitHub)

[Enllaç](#) al fitxer html del codi d'R. (dins del repositori de GitHub)

[Enllaç](#) per descarregar tot el Repositori de GitHub.

Referències

- [1] Generalitat de Catalunya. *dades obertes*. URL: <https://analisi.transparenciacatalunya.cat/>.
- [2] Generalitat de Catalunya (departament d'interior). *Accidents de trànsit amb morts o ferits greus a Catalunya*. URL: <https://analisi.transparenciacatalunya.cat/Transport/Accidents-de-tr-nsit-amb-morts-o-ferits-greus-a-Ca/rmgc-ncpb>.
- [3] David Candela Rubio. *repositori de GitHub*. URL: https://github.com/Littleote/UNI_AdDC_Exam.