

Grau en Estadística

Títol: Factors explicatius de la taxa de criminalitat de les regions espanyoles

Autor: Alba Bernal Aparicio

Director: Jordi Pons i Novell

Departament: Departament d'Econometria, Estadística i Economia Aplicada

Convocatòria: Gener 2024



Agraïments

En especial, al meu tutor Jordi Pons, en la part acadèmica per la seva disponibilitat, entrega i implicació tant resolent dubtes com plantejant millores, i sobretot en la part humana, pel seu acompanyament al llarg de tot el projecte, la seva empatia i el seu inestimable tracte. Com també al meu professor de l'assignatura Models Lineals Generalitzats, Jordi Cortès, qui ha estat molt present en la resolució dels dubtes i inquietuds del meu projecte.

Com també als meus pares, les meves germanes i les meves amigues que han estat un gran recolzament al llarg del projecte, sobretot en la meva estada a Suècia. Els hi dec tot.

Resum

Aquest projecte té com a objectiu comprendre la dinàmica delictiva i determinar els factors que expliquen la taxa de criminalitat de delictes menors i greus a Espanya. S'apliquen models de recomptes amb la introducció de l'offset per a obtenir la taxa de criminalitat. Inicialment, s'ajusta el model de Poisson i, a causa de la sobredispersió, es procedeix amb el model Binomial Negatiu, aplicant transformacions a les variables per a millorar la seva validació. S'identifiquen factors significatius per a cada tipus de delictes. En delictes menors, l'edat de la població, la presència d'estrangers i la pobresa són significatius, mentre que, en delictes greus, l'edat, la pobresa, l'educació superior i el nombre de divorcis són destacats.

Paraules clau: model de recomptes, taxa de criminalitat, delictes menors, delictes greus, odds rati, offset, model Poisson i model Binomial Negatiu.

Classificació AMS (MSC2010):

62J12 Models lineals generalitzats

62H20 Mesures d'associació

62-07 Anàlisi de dades

62-09 Mètodes gràfics

62M10 Sèries temporals, autocorrelació, regressió, etc.

Resumen

Este proyecto tiene como objetivo comprender la dinámica delictiva y determinar los factores que explican la tasa de criminalidad de delitos menores y graves en España. Se aplican modelos de recuentos con la introducción del offset para obtener la tasa de criminalidad. Inicialmente, se ajusta el modelo de Poisson y, debido a la sobredispersión, se procede con el modelo Binomial Negativo, aplicando transformaciones a las variables para mejorar su validación. Se identifican factores significativos para cada tipo de delito. En delitos menores, la edad de la población, la presencia de extranjeros y la pobreza son significativos, mientras que, en delitos graves, la edad, la pobreza, la educación superior y el número de divorcios son destacados.

Palabras clave: modelo de recuentos, tasa de criminalidad, delitos menores, delitos graves, odds ratio, offset, modelo Poisson y modelo Binomial Negativo.

Classificación AMS (MSC2010):

62J12 Modelos lineales generalizados

62H20 Medidas de asociación

62-07 Análisis de datos

62-09 Métodos gráficos

62M10 Series temporales, autocorrelación, regresión, etc.

Abstract

This project aims to understand the dynamics of crime and determine the factors that explain the crime rate of minor and serious crimes in Spain. Counter models are applied with the introduction of offset to obtain the crime rate. Initially, the Poisson model is adjusted and, due to overdispersion, we proceed with the Negative Binomial model, applying transformations to the variables to improve its validation. Significant factors for each type of crime are identified. In minor crimes, the age of the population, the presence of foreigners and poverty are significant, while in serious crimes, age, poverty, higher education and the number of divorces is prominent.

Keywords: count model, crime rate, minor crimes, serious crimes, odds ratio, offset, Poisson model and Negative Binomial model.

AMS Classification (MSC2010):

62J12 Generalized linear models

62H20 Association measures

62-07 Data analysis

62-09 Graphic Methods

62M10 Temporary series, self-correlation, regression, etc.

SUMARI

I.	INTRODUCCIÓ I OBJECTIUS DEL PROJECTE.....	1
II.	METODOLOGIA.....	3
III.	MARC TEÒRIC.....	6
	3.1 Definició del crim.....	6
	3.2 Factors de la criminalitat	7
IV.	ANÀLISI DESCRIPTIVA.....	10
	4.1 Anàlisi temporal i geogràfic	10
	4.1.1 Anàlisi trimestral	13
	4.2 Anàlisi de les tipologies penals	22
	4.3 Conclusions de l'anàlisi descriptiva	31
	4.3.1 Conclusions anàlisi descriptiva trimestral	31
	4.4 Conclusions de l'anàlisi de tipologies penals	31
V.	MODELS ESTADÍSTICS	33
	5.1 Models Log-lineals.....	33
	5.1.1 Model Poisson.....	33
	5.1.2 Model Quasi-Poisson	37
	5.1.3 Model Binomial Negatiu	37
VI.	RESULTATS.....	39
	6.1 Tractament i selecció de variables	39
	6.2 Anàlisi descriptiva bivariant.....	40
	6.2.1 Delictes menors per CCAA	41
	6.2.2 Delictes greus per CCAA.....	41
	6.2.3 Matriu de correlació variables explicatives	42
	6.2.4 Variable indicadora de pandèmia	43
	6.3 Models log-lineals emprats.....	44
	6.3.1 Model Poisson.....	44
	6.3.1 Model Binomial Negatiu	48
VII.	DISCUSSIÓ I CONCLUSIONS.....	56
	BIBLIOGRAFIA	60
	ANNEX	65
	A. Gràfics complementaris	65
	B. Taules complementàries	76
	C. Codi en llenguatge R	76

I. INTRODUCCIÓ I OBJECTIUS DEL PROJECTE

La criminologia és una disciplina multidisciplinària que busca comprendre i analitzar els factors que contribueixen a la realització d'actes delictius en una societat. La detecció de patrons i tendències en la delinqüència és essencial pel disseny de polítiques públiques efectives, l'assignació de recursos i la presa de decisions informades per part de les autoritats. En aquest context, el present treball de fi de grau (TFG) es planteja des del punt de vista de l'estadística aplicada a la criminologia. Se centra en l'anàlisi multivariant d'indicadors en criminologia, amb l'objectiu d'investigar la prevalença de crims a diferents comunitats i ciutats autònomes d'Espanya i, en particular, a identificar els factors crucials que incideixen en l'execució d'actes delictius en aquestes regions.

En una societat en constant evolució, la comprensió dels determinants de la delinqüència és de vital importància. Aquest estudi se centra en una varietat d'indicadors que podrien influir en la taxa de criminalitat a les diferents comunitats i ciutats autònomes. Aquests indicadors inclouen el nivell educatiu de la població, la situació econòmica de les persones (inclòs l'atur), les taxes de separacions i divorcis, així com la presència d'immigrants i turistes a cada regió.

Mitjançant un enfocament d'anàlisi multivariant, es busca explorar la relació complexa i sovint interconnectada entre aquests indicadors i la incidència delictiva. L'anàlisi multivariant permet una comprensió més profunda de com múltiples factors poden contribuir conjuntament als nivells de criminalitat en una àrea geogràfica determinada. En identificar els factors determinants i les seves relacions, aquest estudi pretén proporcionar informació valuosa que pot ser utilitzada pel desenvolupament d'estratègies i polítiques orientades a la prevenció del delictes i la seguretat ciutadana.

A més, el present treball s'estructura en una sèrie de seccions que aborden tant els aspectes teòrics com metodològics necessaris per a dur a terme una anàlisi multivariant dels indicadors en criminologia. Es comença ajustant models de Poisson, com a conseqüència de l'aparició de la sobredispersió s'acaben ajustant models Binomials negatius per a tractar d'explicar els factors que incideixen en la taxa de criminalitat de delictes menors i greus. S'espera que els resultats obtinguts d'aquesta investigació contribueixin a una comprensió més profunda de la delinqüència en el context regional a Espanya, ajudant així en la formulació de polítiques públiques més efectives i en la presa de decisions informades.

La metodologia del projecte es fonamenta en dades públiques oficials proporcionades pel Ministeri d'Interior i l'Institut Nacional d'Estadística (INE). Aquestes dades es converteixen en la matèria prima essencial per a l'anàlisi estadístic, que examina la quantitat de delictes, la

seva distribució temporal i espacial, així com la influència de factors demogràfics, econòmics, educatius i socials.

Pel que fa a les hipòtesis, la primera d'elles es planteja que existeix una variabilitat significativa en les taxes de criminalitat entre les diferents regions. Aquesta suposició suggereix que la incidència delictiva no segueix un patró uniforme a tot el territori espanyol. A més, es planteja la hipòtesi de la diferent evolució temporal de criminalitat amb patrons específics en cada regió.

La tercera hipòtesi se centra en les diferències significatives en la prevalença dels diferents tipus de delictes entre les diferents comunitats i ciutats autònomes. Aquesta suposició reflecteix la possibilitat que factors socials, econòmics o culturals específics de cada regió puguin influir en la dinàmica delictiva i determinar la seva composició.

Finalment, la quarta i última hipòtesi destaca la relació entre els factors demogràfics, econòmics, educatius i socials seleccionats amb la taxa de criminalitat. A priori, s'explora la possibilitat que una menor educació es pugui relacionar amb una major incidència en l'execució de delictes, suggerint que el nivell d'instrucció pot ser un factor contribuent a la dinàmica criminal. A més, es planteja la idea que una situació econòmica desafavorida podria incentivar la comissió de delictes, posant de manifest la possible relació entre les condicions econòmiques i la conducta delictiva.

Pel que fa a l'estat civil de la persona, els divorcis podrien ser un motiu a tenir en compte per l'execució dels crims passionals. Aquesta associació suggereix que les tensions i les emocions vinculades als processos de separació podrien contribuir a la comissió de cert tipus de delictes entre les parelles afectades i els seus descendents.

En referència al turisme, es considera que pot tenir una relació amb alguns tipus de delictes, probablement influït per factors com l'afluència de persones, la diversitat cultural i altres elements que podrien afectar la dinàmica del comportament delictiu en aquest context específic.

A través d'aquestes hipòtesis, es busca comprendre i contextualitzar la dinàmica delictiva a nivell regional a Espanya.

II. METODOLOGIA

L'objectiu de la investigació és comprendre la dinàmica delictiva i determinar quins factors afecten a la taxa de criminalitat. Com la variable d'interès és el nombre total de delictes – que amb l'afegiment de l'offset esdevindrà la taxa de criminalitat (nombre de crims convencionals per a cada 1.000 persones) – cal aplicar un model de recomptes. El model estadístic que cal triar, d'entrada, és el de Poisson, que assumeix que la variància de les observacions és igual a la seva mitjana. Després s'haurà d'estudiar la sobredispersió com també la presència de 0 per adaptar-ho a un model més adient.

L'offset és una característica important en els models de recompte i es fa servir quan es vol ajustar el model a dades amb diferents temps d'exposició o diferents grandàries de població. L'offset és una variable amb un coeficient fix d'1. Això significa que es tracta d'una variable amb un pes constant en el model, la qual cosa permet comparar directament les taxes en lloc de simplement les freqüències absolutes. Com hi ha diferents dimensions poblacionals o àrees amb temps d'exposició variables, resulta útil afegir un offset per ajustar les diferències en la grandària de població o àrea. L'ús de l'offset amb el logaritme de la població (o la mida de l'àrea) assegurarà que s'està modelant la taxa de criminalitat per unitat de població, permetent comparacions significatives entre àrees o temps.

Respecte la procedència de les dades, s'usen les dades públiques del Ministeri d'Interior¹ i de l'Institut Nacional d'Estadística (INE). Les estadístiques públiques del Ministeri d'Interior provenen de les forces policials, la fiscalia general de l'Estat, els tribunals de justícia i la direcció general de institucions penitenciàries. Aquestes estadístiques provenen de dos sistemes diferents: el Sistema Estadístic de Criminalitat (SEC) des del gener del 2015 fins a l'actualitat i el Programa Estadístic de Seguretat (PES) per a dates anteriors, utilitzant totes les dades disponibles. En el cas de les dades anuals, s'usen els anys 2010-2022. Pel cas de les trimestrals des del 2015 fins el segon trimestre del 2023.

El treball de final de màster de l'Àngel González (2021)² [15] ha permès obtenir els indicis de quins factors podrien afectar a la taxa de criminalitat convencional. Amb un petit canvi en la variable resposta, separant els crims convencionals dels que afecten directament a la persona perquè els factors influeixen de diferent manera segons Morales M. (2017) [26]. Les variables que s'usen en aquest estudi per a cada ciutat autònoma són les següents:

¹ Per a un major detall consultar la pàgina web:

<https://estadisticasdecriminalidad.ses.mir.es/publico/portalestadistico/>

² Per a un major detall consultar la pàgina web: <https://zagan.unizar.es/record/109787>

Taula 13. Variables usades en el projecte.

Categoria	Variables
Tipologies Penals	Homicidis dolorosos i assassinats consumats Homicidis dolorosos i assassinats en grau de temptativa Delictes greus i menys greus de lesions i baralla tumultuària Segrests Agressió sexual amb penetració Resta de delictes contra la llibertat i indemnitat sexual Robatoris amb violència i intimidació Robatoris amb força en domicilis, establiments i altres instal·lacions Furts Sostraccions de vehicles Tràfic de drogues Resta de criminalitat convencional Criminalitat convencional total
Variable Anual	Nombre total de criminalitat convencional Delictes contra la persona
Variables Demogràfiques	Població subjecta a lleis penals Taxa d'envelliment de la població Nombre de persones estrangeres que viuen a Espanya Nombre de turistes estrangers
Variables Econòmiques	PIB per càpita Renda mitjana de les llars Taxa de risc de pobresa Nombre d'aturats en milers de persones
Variables Educatives	Taxa d'abandonament prematur dels estudis Percentatge de persones amb nacionalitat espanyola d'entre 25-64 anys que tenen el nivell d'estudis inferior a la 2a etapa d'Educació Secundària Percentatge de persones amb nacionalitat espanyola d'entre 25-64 anys que disposen de la 2a etapa d'Educació Secundària Percentatge de persones amb nacionalitat espanyola d'entre 25-64 anys que disposen d'Educació Superior
Variable Social	Nombre de divorcis per a cada mil habitants

Font: Elaboració pròpia.

És important reconèixer la existència de biaix, com també la presència de dades fosques en les estadístiques oficials sobre la delinqüència. Tradicionalment, aquestes estadístiques han estat objecte de crítiques degut a les seves limitacions com a mesures precises i completes del delictes real en les societats occidentals contemporànies. Molts criminòlegs argumenten que aquestes estadístiques reflecteixen una construcció social esbiaixada del fenomen del crim basada en denúncies oficials. Sostenen que les estadístiques són incompletes ja que només consideren conductes il·lícites denunciades per les víctimes a la policia. Però, donat que no totes les víctimes denuncien, es registra només una part limitada de tots els comportaments delictius reals, contribuint al que es coneix com la xifra fosca del delictes (Vozmediano & San Juan, 2010) [37]. A més, els registres presenten biaixos per dues raons. En primer lloc, alguns

delictes són més fàcils de denunciar o detectar que altres, el que augmenta la probabilitat que apareguin als registres oficials. En segon lloc, existeix una dificultat tècnica per a traduir certes situacions socials a categories estadístiques específiques, i es qüestiona la consistència en el procés de denuncia i complicació estadística (M. Ruiz, 2018) [31].

Malgrat les crítiques, els recomptes de delinqüència proporcionats pel Ministeri d'Interior són els únics indicadors oficials i de validesa multiescalar per a mesurar l'impacte territorial del delicte a Espanya. En aquest treball, tenint en compte les seves limitacions, s'han usat aquestes estadístiques per a aproximar-se al model evolutiu i distribucional del crim al territori espanyol.

Els recursos informàtics usats han estat essencialment l'Excel per a la neteja de les bases de dades i el càlcul de la taxa de criminalitat, i l'R pels anàlisis descriptius univariants i bivariants, com també per l'avaluació i validació del model.

III. MARC TEÒRIC

L'àmbit de la criminologia és molt extens, es tracta de la ciència que estudia tant la delinqüència individual com la criminalitat en general. Es tracta d'una ciència multidisciplinària, ja que el crim es troba present en nombrosos àmbits (socials, psicològics, jurídics...). D'aquesta manera, es necessiten conèixer les causes, els factors o condicions que fan que es produeixi el crim.

Segons Teresa del Socorro, en el seu llibre titulat *L'enfoc psicològic en la criminalista* (2021) [32], s'identifiquen 3 nivells d'enfoc en el camp de la Criminologia, interrelacionats entre si. El primer, és el conductual, que es dedica a l'estudi detallat del comportament antisocial a nivell individual. El segon, és el personal, que busca comprendre la personalitat del delinqüent amb la finalitat de facilitar el seu tractament i rehabilitació. El tercer, de caràcter general, es fonamenta en l'anàlisi integral de les conductes delictives i les seves característiques en un lloc i temps específics. Aquest tercer nivell d'interpretació requereix l'ús d'una metodologia estadística per a descobrir patrons o característiques típiques en aparents irregularitats, establir relacions entre diversos fenòmens i deduir les lleis generals que els regeixen. No obstant, cal tenir cura de l'ús d'aquesta metodologia, ja que la seva precisió està fortament influenciada per la terminologia emprada, la tipificació legal, els criteris polítics, d'entre altres factors. Per això, resulta essencial tenir un coneixement profund de les definicions dels conceptes relacionats amb les dades de l'estudi.

3.1 Definició del crim

En aquesta recerca, la criminalitat s'entén com un conjunt d'activitats i de comportaments que són considerats delictius o il·legals per la societat o per la llei (E. García, 2019) [13]. Es tracta d'un fenomen social que aborda una ampla gamma de comportaments, des dels delictes menors fins als greus. És, per tant, la societat qui determina quines conductes són les permeses.

El delictes és una manifestació concreta de la criminalitat i es defineix com una acció o omissió que està prohibida i castigada per la llei pertinent. Per a que existeixi un delictes, la conducta ha de ser tipificada com condemnable en el Codi Penal i ha de complir amb els elements de ser típica, antijurídica i culpable. La delimitació del que constitueix un delictes pot variar entre els diferents països i al llarg del temps, ja que està influenciada pels factors socials, culturals, polítics i econòmics. Un exemple que il·lustra la variabilitat en la definició del delictes és el cas del matrimoni infantil, que és possible en certs països (delictes culturalment condicionat), malgrat ser una violació dels drets humans. Això, a més, pot reflectir les diferències profundes

en les percepcions sobre la moralitat i els drets humans, com també, que la definició i la gravetat del delictes estan subjectes a interpretacions variables en diferents contextos.

Des del punt de vista jurídic, el crim i el delictes són conceptes equivalents, ja que ambdós impliquen conductes punibles. Però, el terme crim s'usa per a referir-se a delictes d'extrema gravetat, delictes particularment greus, especialment aquells que són ofensius contra les persones, com l'assassinat o la tortura.

Encara que els termes delictes i infracció penal comparteixen la característica de referir-se a accions prohibides per la llei, existeixen matisos conceptuals que els diferencien. En l'àmbit legal, el delictes acostuma a estar definit en els codis penals, mentre que la infracció penal pot incloure violacions de normatives que no necessàriament estan estipulades com a delictes penals. A grans trets, el delictes aborda un espectre més ampli de conductes il·lícites, mentre que la infracció penal s'usa de manera més específica per denotar violacions de la llei penal.

D'altra banda, com s'havia avançat anteriorment, és important tenir present que la interpretació del que es constitueix un delictes pot canviar amb l'evolució de la societat i els seus valors. L'aparició de noves tecnologies, com ara l'ús generalitzat de l'ordinador i l'expansió de la internet, ha donat lloc a nous reptes legals i noves formes de delinqüència. Un exemple rellevant d'això són els ciberatacs, una categoria de delictes que ha sorgit amb la digitalització de la societat. La resposta legal a aquesta nova forma de delinqüència ha requerit l'adaptació i el desenvolupament de legislació específica per a abordar els desafiaments únics que planteja l'àmbit cibernètic.

3.2 Factors de la criminalitat

La gran preocupació actual pels factors polítics, econòmics, demogràfics i socio-culturals de la criminalitat, ha conduït a estudiar el fenomen com a un conjunt. La investigació sobre els motius darrera dels comportaments contraproductius ha estat un camp extens i multidisciplinari, buscant identificar factors i condicions que influeixen en la conducta antisocial³. Han estat nombrosos autors els que han estudiat les causes i els factors: elements demogràfics com la immigració (Fernández Bessa, 2010) [16], factors econòmics com la pobresa (Heller, Jacob & Ludwig, 2011) [18], elements de l'àmbit educatiu (Cáceres, Farro, Uribe & Chamorro, 2022) [4] i aspectes de la vida familiar com els divorcis (César Pol & Silvestrini, 2004) [8].

³ Patrons de comportament que van en contra de les normes socials i que, amb freqüència, infringeixen els drets dels demés, pertorbant el ordre social (Sánchez Arteaga, 2021) [32].

Pel que fa a l'àmbit demogràfic, les variables d'immigració i turisme han estat molt analitzades i controvertides. En el nostre dia a dia, podem escoltar moltes frases xenòfobes en contra de la immigració tal com: “venen a delinquir i no s'adapten”, sovint recolzades en discursos d'odi en mitjans de comunicació i en algunes entitats polítiques al llarg dels anys. No obstant, existeixen nombroses investigacions que indiquen just el contrari, que la immigració no augmenta els índexs de criminalitat (E. García, 2019) [13]. De fet, existeixen característiques dels estrangers que contribueixen a que les taxes no incrementin (Asociados, 2022) [2]. A més tenen impactes positius en el país de destí, segons l'informe WEO del 2020 del Fons Monetari Internacional (FMI), la immigració es tradueix en un creixement econòmic en quant al producte i la productivitat a curt i mitjà termini en països avançats en tecnologia.

La relació entre el turisme i la criminalitat ha estat objecte de diversos estudis amb resultats contradictoris. Alguns experts, com David Navarrete (2014) [27] i Luis David (2014) [29], suggereixen que la quantitat de turistes en un territori amb prou feines té correlació amb els índexs de delinqüència. Malgrat això, investigacions científiques pel territori espanyol, com les de Montolio i Planells (2012) [25], contradiuen aquesta visió al determinar que si existeix una relació, especialment, en els delictes comesos contra les persones i el patrimoni. Aquest fet destaca la complexitat de la dinàmica entre el turisme i la criminalitat, mostrant que les conclusions poden variar segons la regió o inclús per l'enfoc específic de l'estudi.

L'àmbit econòmic es revela com un factor central en la comprensió de la criminalitat. Alguns estudis suggereixen que la relació entre els ingressos i la delinqüència no és uniforme per a tots els delictes, essent més evident en aquells contra la propietat, segons indiquen Corman & Mocan (2005) [10] i Gerstenblith (2016) [17]. Altres autors, com Natalia Estévez (2021) [12], argumenten que la criminalitat tendeix a augmentar amb la disminució d'ingressos, especialment en situacions d'atur involuntari. Aquest vincle destaca encara més en delictes contra la propietat, segons assenyalen Julie Phillips i Kenneth C. Land (2012) [28] i Natalia Estévez (2021) [12]. Malgrat, Phillips i Land, adverteix que la connexió entre el delicte i l'atur és dèbil i que està fortament correlacionada amb altres variables de mercat, generant resultats variables.

Per últim, en aquest context econòmic, la taxa de pobresa també sorgeix com un tema recurrent en la discussió sobre la delinqüència. El 13è congrés de les Nacions Unides sobre la Prevenció del Delicte i Tractament Delinqüent en el 2015 va concloure que la condició de pobresa estava present en les zones amb major índexs delictius. A més, investigacions detallades, com les de Téllez & Soler (2022) [33], estableixen que la carència i marginalitat associades a la pobresa afecten significativament en la criminalitat. D'aquesta manera, la

comprensió dels factors econòmics es torna essencial per a abordar la complexitat dels índexs delictius.

En abordar els agents causants o afavoridors del delictes, és imperatiu examinar la perspectiva educacional dels individus en una població. S'ha argumentat que el delictes està vinculat a conjunts poblacionals amb un menor nivell d'escolarització. Globalment, s'observa que un dèficit educatiu contribueix al augment dels delictes, de forma general, s'associa un major nivell d'educació amb una menor probabilitat de cometre actes delictius. En una investigació científica de Carvajal i Rincón (2016) [7], es destaca que una escolaritat alta té un efecte mitigador en les conductes antisocials. Per contra, el fracàs escolar, sovint, condueix a problemes emocionals i d'autoestima, generant conductes disfuncionals i/o delictives, com assenyala Cardona (2020) [5]. És important adonar-se de que la deserció escolar no es considera com un factor de risc aïllat, sinó que opera juntament amb altres elements com la falta d'assistència a classe i relacions problemàtiques entre iguals, configurant-se com la "primera fita a la trajectòria delictiva", segons Barreto (2002) [3]. A més, estudis, com el realitzat per Machin, Marie i Vujic (2011) [24], demostren que el establiment de l'educació secundària obligatòria, o qualsevol mesura que allargui la permanència en les institucions acadèmiques, condueix a reduccions significatives en la comissió de delictes. Investigadors, com Lochner (2008) [23], sostenen que les polítiques de prevenció del delictes a nivell estatal han de centrar-se en la educació, ja que la instrucció civilitza i educa el comportament dels individus.

Malgrat la evidència demostrada per molts investigadors de la relació positiva entre educació i prevenció del crim, cal assenyalar que un major nivell educatiu no té efectes preventius en els delictes de coll blanc⁴ i, en certes circumstàncies, podria inclús augmentar-los, segons les investigacions de Lochner (2008) [23]. En el context espanyol, els estudis sobre el crim han abordat l'educació des de dades agregades, considerant-la com un possible element limitador del delictes dins d'un conglomerat major en el que altres determinants també tenen un paper rellevant.

Finalment, també existeixen els factors de risc en l'àmbit social. Es relacionen, generalment, amb poblacions o personalitats vulnerables que tenen una major propensió a conductes delictives. Les xarxes de recolzament poden influir significativament en el creixement i la formació d'un individu, i la seva absència podria donar lloc al fracàs escolar, com també a augmentar la probabilitat de conductes delictives (Cáceres, Farro, Uribe & Chamorro, 2022)

⁴ Delictes financers, econòmics o comercials comesos per persones d'alt estatus social, sovint en el context de les seves ocupacions professionals. (Gerstenblith, 2016) [17].

[4]. D'altra banda, la relació de la trajectòria delictiva amb la vulnerabilitat i la exclusió social ha estat destacada, evidenciant que a major vulnerabilitat i exclusió social, augmenta la probabilitat d'involucrar-se en la delinqüència juvenil i de profunditzar en la trajectòria delictiva (Uceda & Domínguez, 2017) [34].

IV. ANÀLISI DESCRIPTIVA

L'anàlisi descriptiva es divideix en dos apartats. El primer aborda l'estudi de l'evolució de la taxa de criminalitat anual (des del 2010 fins el 2022) i trimestral (des del 2015 fins el segon trimestre del 2023) a nivell nacional. El segon apartat recull l'anàlisi quantitatiu de les dotze tipologies delictives, és a dir, quin delicte té una major incidència a nivell nacional i en cada CCAA i ciutat autònoma.

Cal esmentar que les dues ciutats autònomes, Ceuta i Melilla, presenten les seves pròpies característiques. S'ubiquen en la costa del nord d'Àfrica, són regions petites amb població relativament petita, presenten un caràcter fronterer com també major diversitat cultural. Aquestes diferències juntament amb el seu estatus especial i reptes únics en migració, suggereixen que poden comportar-se de manera diferent respecte la resta de comunitats autònomes peninsulars en ajustar el model que inclou totes les regions del territori espanyol. Per tant, considerar aquestes particularitats és crucial per a capturar de manera precisa la seva dinàmica socioeconòmica.

4.1 Anàlisi temporal i geogràfic

Primer de tot, es comenta l'evolució de la taxa de criminalitat anual (número d'infraccions penals per 1.000 habitants) a nivell nacional. Més tard, s'entrarà en major detall en les diferències geogràfiques i temporals de la mateixa taxa.

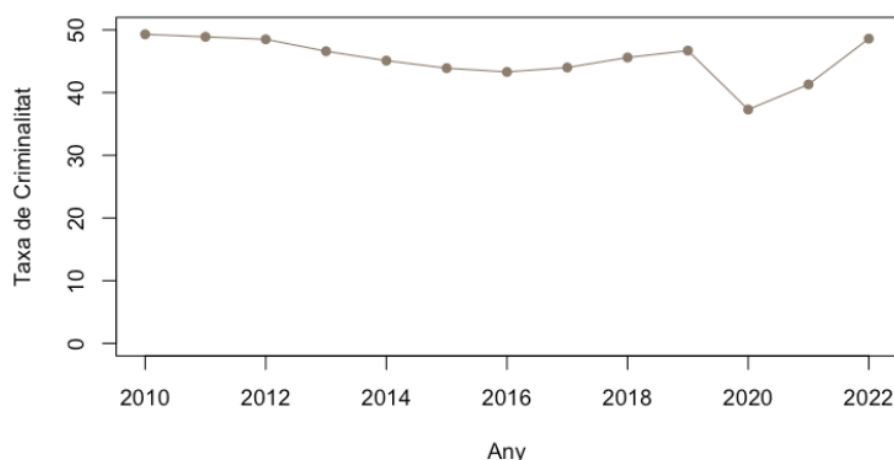
L'anàlisi de l'evolució temporal de la taxa de criminalitat anual durant els anys 2010-2022 revela tendències notables. A primera vista, s'observa, sense tenir en compte l'efecte de la pandèmia global, un comportament més o menys homogeni.

Iniciant-se l'any 2010, amb la major taxa de criminalitat del període estudiat. Durant els anys 2011-2016, es noten lleugeres reduccions en la taxa de criminalitat, indicant una possible millora en la seguretat nacional. Després, al 2017-2018, experimenta un lleuger augment, que podria relacionar-se amb la complexa dinàmica social i econòmica experimentada durant la fase de recuperació econòmica, d'això ja s'aprofundeix més endavant.

L'arribada de la pandèmia, la COVID-19, i les mesures de confinament implementades en el 2020 provoquen una dràstica caiguda en la taxa de criminalitat, marcant una fita singular en aquest anàlisi. Sugerint que les restriccions i el canvi en la dinàmica social durant la pandèmia van tenir un impacte significatiu en l'activitat delictiva.

A partir del 2021, l'increment de la taxa de criminalitat es fa més pronunciat, comparat amb els anys anteriors, revelant una tendència creixent que augmenta en cada període anual successiu, apropant-se als valors de criminalitat del 2010.

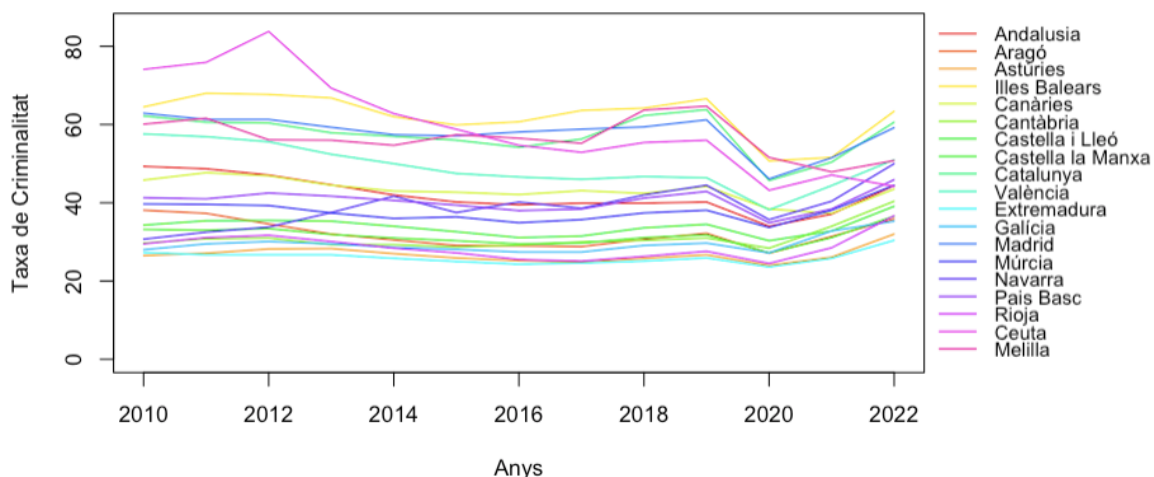
Gràfic 1. Evolució Temporal de la Taxa de Criminalitat nacional anual (2010-2022)



Font: Elaboració pròpia a partir de l'R i les dades de l'INE.

En el següent gràfic es mostra l'evolució temporal de la taxa de criminalitat per cada CCAA durant el període 2010-2022, permetent identificar quines són les zones del país amb menor i major taxa de criminalitat.

Gràfic 2. Evolució Temporal de la Taxa de Criminalitat per CCAA (2010-2022)



Font: Elaboració pròpia a partir de l'R i les dades del Ministeri d'Interior i de l'INE.

Pel que fa a les comunitats i ciutats autònomes, durant el període inicial, Ceuta lidera el rànquing com la regió amb la taxa més elevada, seguida de prop per les Illes Balears fins el 2014, moment en el qual les Illes Balears experimenten la major taxa de criminalitat fins el 2022. Tot i que, només en el 2020, torna a ser superada per Ceuta. A partir del 2015, Madrid, Catalunya i Melilla experimenten un augment gradual en les seves taxes, essent Melilla la que destaca per augmentar les distàncies a partir del 2017. Pel 2022, les regions amb majors taxes són, en ordre creixent, Ceuta, Madrid, Catalunya i les Illes Balears. Per contra, les que presenten taxes més baixes són Extremadura, Astúries i La Rioja.

Cal afegir que s'observen comportaments anòmals, particularment a Ceuta. Entre el 2011 i el 2012, aquesta ciutat autònoma experimenta un augment en la seva taxa, mentre la resta es mantenen estables. Mentre que, en el període 2021-2022, les altres regions augmenten la seva taxa, Ceuta és l'única que mostra una disminució de la mateixa. Aquesta reducció podria donar-se a factors locals, canvis en polítiques de seguretat o inclús per impactes econòmics de la pandèmia. Un altre comportament anòmal a tenir en compte és el cas de Navarra que presenta un pic al 2014, mentre la resta de CCAA experimenten descensos. Aquest comportament atípic podria estar relacionat amb esdeveniments específics locals que s'analitzen més endavant.

A més, la gràfica adjunta, Gràfic 2, ens ofereix una visió dels factors subjacents a les diferències en les taxes de criminalitat entre les CCAA i les ciutats autònomes. Madrid, Catalunya i les Illes Balears destaquen per la condició de ser destins turístics importants, una alta densitat poblacional i una urbanització significativa. Aquestes característiques podrien contribuir a una dinàmica de major incidència delictiva, especialment en àrees turístiques o moments de major afluència. Paral·lelament, les CCAA amb taxes més baixes, com Extremadura, Astúries i La Rioja, mostren una menor densitat poblacional, menor urbanització i una activitat econòmica menys intensa. La homogeneïtat demogràfica en aquestes àrees pot contribuir a una cohesió social que es reflecteix en el bon resultat de les taxes de criminalitat.

Aquestes tendències, tot i tractar-se de generalitzacions, revelen una realitat complexa i multifacètica. És important assenyalar que aquestes generalitzacions no captiven la totalitat del panorama, ja que la incidència de la criminalitat es veu afectada per una àmplia gamma de factors interrelacionats. És necessari fer un anàlisi més detallat i específic per a comprendre plenament les característiques i dinàmiques que influeixen en les taxes de criminalitat.

4.1.1 Anàlisi trimestral

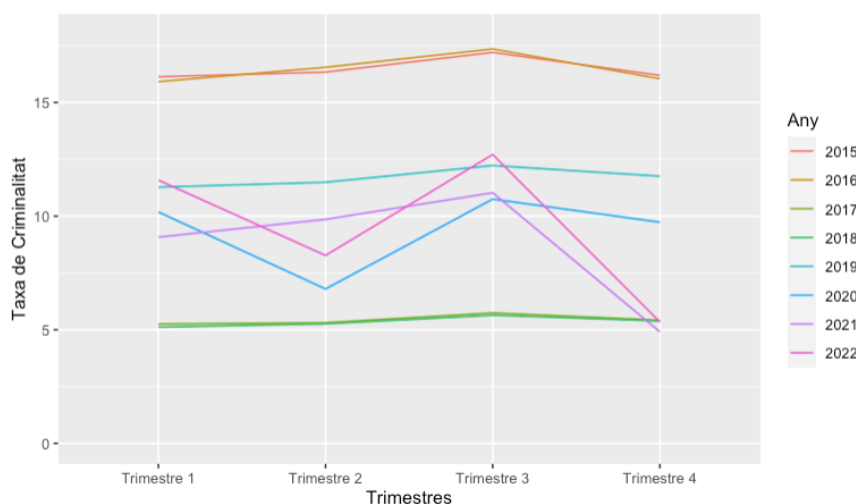
A partir dels balanços trimestrals de criminalitat per CCAA, facilitat pel portal estadístic de criminalitat del Ministeri d'Interior, es procedeix a realitzar un anàlisi de les taxes de criminalitat trimestrals. S'usen totes les dades disponibles a dia d'avui, 20/11/2023, des del primer trimestre del 2015 fins el segon del 2023.

La necessitat de calcular la taxa de criminalitat per a cada comunitat i ciutat autònoma rau en l'objectiu de fer que les dades siguin comparables entre les diferents regions. Per aconseguir aquest objectiu, s'han buscat dades demogràfiques trimestrals per a cada comunitat autònoma. Tot i que, ocasionalment, no disposem de dades pels segons i quarts trimestres, s'ha realitzat una interpolació lineal per estimar-les. Aquesta interpolació lineal ens permet predir valors aproximats en base a les dades disponibles i considerar els canvis en la població al llarg de l'any de forma lineal. Per a més detall sobre la interpolació lineal efectuada, veure als *Annexos* els gràfics i el codi usat en R, a l'apartat de gràfics complementaris i codi d'R.

Cal afegir que, amb l'objectiu de mantenir la integritat dels resultats i evitar inferències inadequades, no s'ha realitzat interpolació en les dades corresponents al segon, tercer i quart trimestres del 2023. Aquesta decisió es basa en el caràcter provisional d'aquestes dades, amb l'enteniment que no s'ha buscat efectuar prediccions, sinó mostrar l'evolució real de la taxa de criminalitat fins a la data disponible. És per aquest motiu que en els gràfics es mostrarà l'evolució trimestral de la taxa de criminalitat des del primer trimestre del 2015 fins l'últim del 2022.

El Gràfic 3, que presenta l'evolució trimestral de la Taxa de Criminalitat a Espanya des del 2015 al 2022, revela patrons distintius en diferents períodes.

Gràfic 3. Evolució de la taxa de criminalitat trimestral a Espanya. 2015-2022.



Font: Elaboració pròpia a partir de l'R i les dades del Ministeri d'Interior i de l'INE.

En primer lloc, destaca un augment generalitzat en la taxa de criminalitat en el tercer trimestre de cada any, essent més pronunciat aquest augment en els anys 2020 i 2022, i amb increments menys notables durant els anys 2017 i 2018. Per contra, el quart trimestre destaca per una reducció generalitzada de les taxes de criminalitat. Ara bé, en diferents magnituds, ho fa de forma dràstica pels anys 2021 i 2022, i de forma més subtil per la resta dels anys, sobretot pel 2017 i 2018.

En segon lloc, a l'hora d'examinar els màxims i mínims trimestrals, es destaca que el tercer trimestre registra consistentment les taxes més elevades de criminalitat per a tots els anys estudiats. De fet, és en el tercer trimestre del 2016 on es troba el màxim absolut, amb un valor de 17,36 delictes comesos per cada 1.000 persones. Per contra, del 2015 al 2019, les taxes més baixes es troben en el primer trimestre, mentre que, del 2020 al 2022, les taxes més baixes es traslladen al quart trimestre. Ara bé, el mínim absolut es troba al quart trimestre del 2021, de 4,91 delictes comesos per cada 1.000 persones.

En tercer lloc, s'observen anomalies en les tendències dels anys, revelant una variabilitat significativa en l'evolució de la taxa de criminalitat trimestral a Espanya. La primera és que, pel cas del segon trimestre, cada any té un comportament diferent, n'hi ha que augmenten la seva taxa com ara el 2016 i el 2021 i d'altres que la redueixen com és el cas dels anys 2020 i 2022. La segona té a veure amb la presència de similituds notables en la tendència i la magnitud de la taxa de criminalitat entre el 2015 i el 2016 per una banda, i el 2017 i 2018 per l'altra. Mentre que, pels anys 2020, 2021 i 2022, es fan notables les dràstiques reduccions, marcant una divergència clara respecte els altres anys, que experimenten reduccions més lleugeres. Aquest contrast en la velocitat d'augment i disminució de les taxes durant els anys estudiats destaca la complexitat dels factors socioeconòmics que influeixen en la criminalitat, alhora que subratlla la importància de considerar contextos específics o esdeveniments extraordinaris, com ara la pandèmia de la COVID-19, que puguin haver contribuït a aquests patrons.

Per a l'anàlisi detallat de l'evolució de la taxa de criminalitat trimestral en cada comunitat i ciutat autònoma, es comença comentant aquelles regions que mostren les majors taxes. A continuació, s'exploraran algun cas que presenti un comportament singular, brindant una comprensió més profunda de les dinàmiques subjacents en aquella regió específica. Finalment, es tancarà l'anàlisi amb les comunitats autònomes més segures. Aquest enfoc permet identificar patrons específics i tendències que poden diferir entre les diferents regions espanyoles per factors locals, demogràfics o econòmics pertinents.

Pel cas català, durant el tercer trimestre, es registra un augment generalitzat, mentre que en el quart trimestre s'observen reduccions generalitzades. D'altra banda, el segon trimestre presenta varies anomalies, ja que cada any es comporta d'una manera diferent, de vegades hi ha augments, com és el cas del 2016 i 2021, i d'altres vegades, es mostren reduccions (2020 i 2022). La menor taxa observada es troba en el quart trimestre del 2021 (7,22 delictes per 1.000 persones) i la major en el tercer trimestre del 2016 (23,51 delictes per 1.000 persones).

Pel cas de les Illes Balears, s'observa un patró similar al cas català, es mostra un augment generalitzat en el tercer trimestre de la taxa de criminalitat i reduccions generalitzades en el quart trimestre. El segon trimestre presenta una gran diferència pel any 2020 que és la única taxa que es veu reduïda. Aquesta diferència podria explicar-se sota els efectes de la COVID-19. La menor taxa es registra en el primer trimestre del 2018 (2,09 delictes per 1.000 persones) i la major en el tercer trimestre del 2015 (35,94 delictes per 1.000 persones).

En contrast, Madrid no segueix un patró consistent al llarg dels trimestres i els anys. Els anys amb comportaments atípics són el 2020 i el 2022 amb elevats ascensos i descensos entre el segon i quart trimestre. L'any 2021 també sorprèn en el quart trimestre per la gran reducció que presenta. La menor taxa es troba en el quart trimestre del 2021 (6,61 delictes per 1.000 persones), mentre que la major es registra en el quart trimestre del 2016 (23,91 delictes per 1.000 persones).

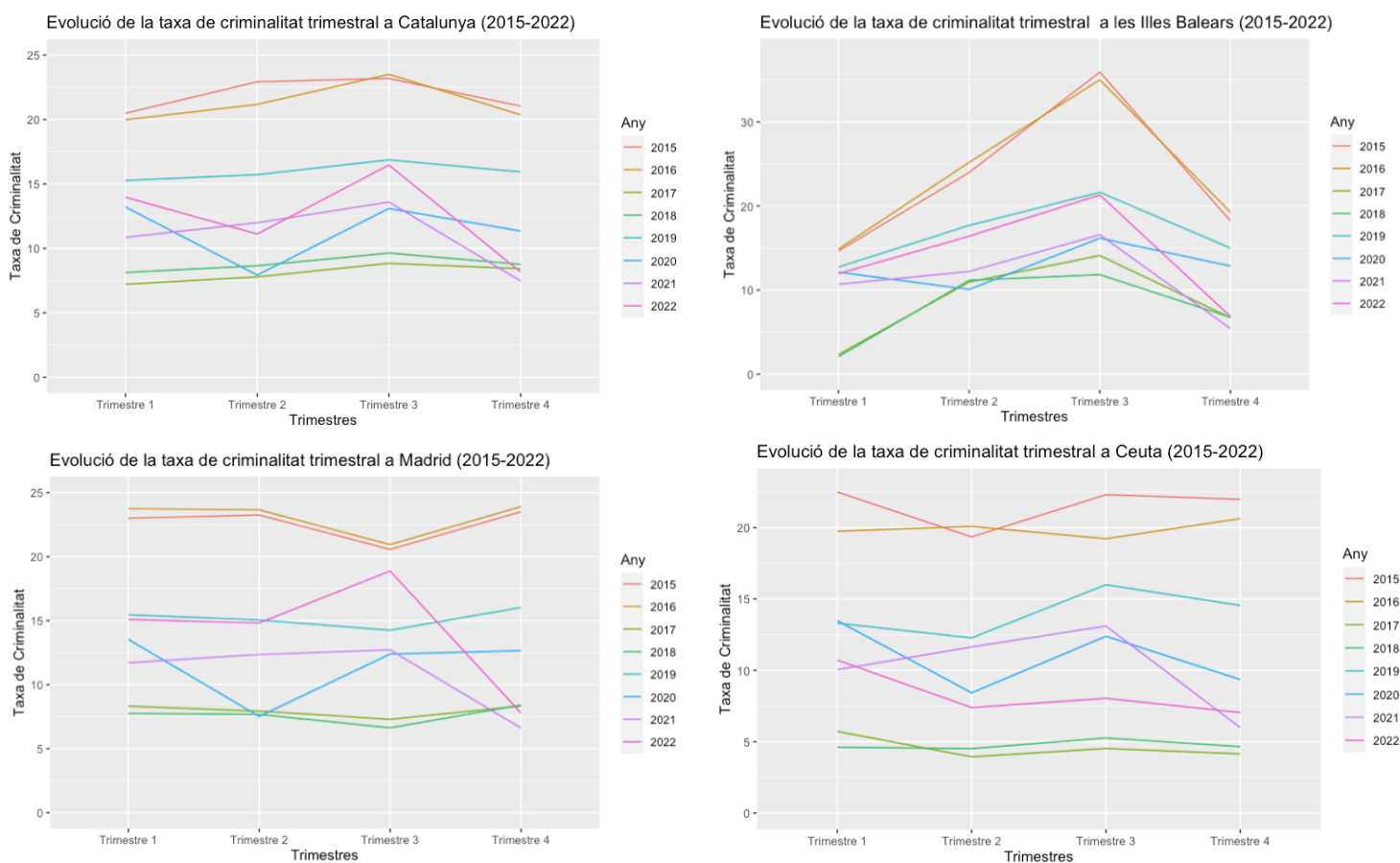
A la ciutat autònoma de Ceuta, l'evolució trimestral de la taxa de criminalitat mostra variacions segons l'any. En el segon trimestre, s'observen increments, excepte pels anys 2015, 2017, 2019 i 2020. Al tercer trimestre també s'observen augments, excepte per l'any 2016. El quart trimestre només presenta ascensos pel 2016. El mínim absolut és del segon trimestre del 2017 amb un valor de 3,94 delictes per 1.000 persones i el màxim és de 22,50 delictes per persones en el primer trimestre del 2015.

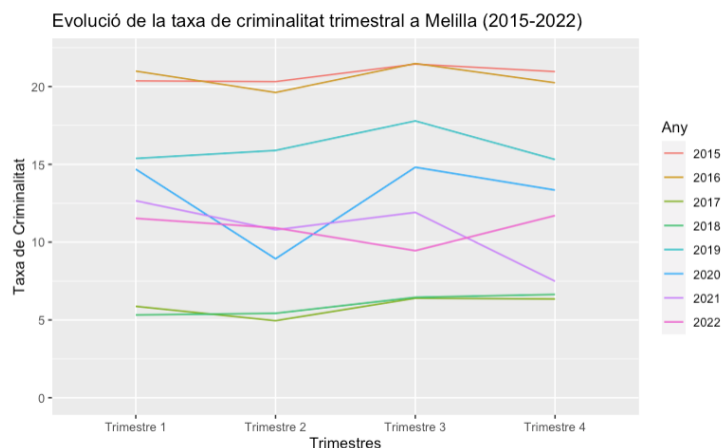
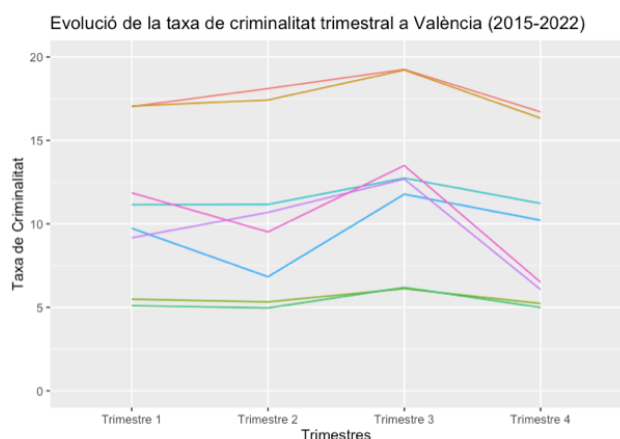
Pel cas valencià, s'observen ascensos del primer al segon trimestre en la majoria dels anys, essent més notables en alguns casos, excepte en el 2022 i el 2020 on es redueix la taxa. De fet, aquests dos últims anys presenten comportaments semblants fins el tercer trimestre, es pot observar en el gràfic adjunt que les seves línies són paral·leles. En el tercer trimestre, augmenten les taxes, alguns anys d'una manera més intensa, mentre que en el quart trimestre les taxes es veuen reduïdes. El mínim absolut es troba en el quart trimestre del 2018 (4,96 delictes per 1.000 persones), i el màxim, en el tercer trimestre del 2015 (19,26 delictes per 1.000 persones).

Finalment, a Melilla, l'altra ciutat autònoma, presenta descensos generalitzats del primer al segon trimestre, menys pels anys 2015, 2018 i 2019. En el tercer trimestre, tots els anys, excepte el 2022, experimenten augments en la taxa de criminalitat. Pel contrari, en el quart trimestre, tots els anys, excepte el 2022, mostra reduccions en la taxa. S'identifica el mínim absolut en el segon trimestre del 2017 (4,95) i el màxim en el tercer trimestre del 2016 (21,48).

Dels casos analitzats, les regions amb majors taxes de criminalitat, es destaca una variabilitat significativa en l'evolució trimestral de les taxes de criminalitat en les diferents regions. Malgrat les fluctuacions, s'observa una tendència generalitzada d'augment durant el tercer trimestre, seguit de reduccions en el quart trimestre en la majoria de les àrees analitzades anteriorment. A més, el segon trimestre presenta comportaments diversos cada any, evidenciant la influència de factors específics. Cal destacar que, al llarg del temps, s'han experimentat millores en les taxes de criminalitat en aquestes regions, ja que els màxims absoluts acostumen a pertànyer als primers anys del període estudiat (2015, 2016).

Gràfic 4. Conjunt de gràfics de l'evolució de la taxa de criminalitat trimestral a Catalunya, Illes Balears, Madrid, Ceuta, València i Melilla. 2015-2022.



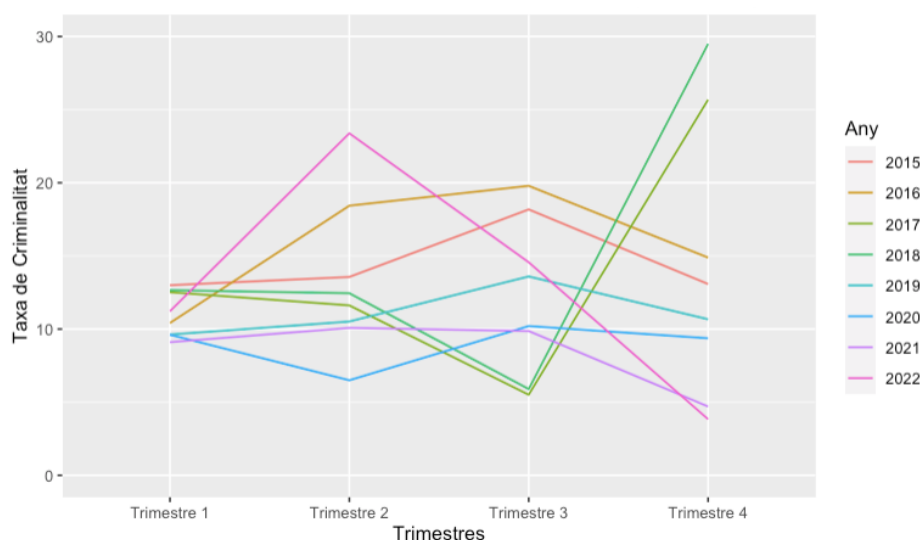


Font: Elaboració pròpia a partir de l'R i les dades del Ministeri d'Interior i de l'INE.

A continuació es procedeix a examinar Navarra, Castella la Manxa i el País Basc, ja que aquestes CCAA són exclusives en exhibir augments significatius en les taxes de criminalitat del 2022, com també evidencien diferents comportaments respecte la resta de regions analitzades.

Pel cas de Navarra, és clau examinar l'evolució de la taxa de criminalitat de forma individual degut al seu comportament temporal. Hi ha una falta d'un patró per a cada trimestre, ja que no queda clar si, depenent del trimestre, la taxa s'incrementa o es redueix. Com a conseqüència trobem el màxim i mínim en el mateix trimestre, el màxim absolut pertany al 2018 (29,49) i el mínim al 2022 (3,84) del quart trimestre. També s'observen anomalies entre els anys 2018 i 2017 respecte els altres anys, ja que quan la taxa s'incrementa en aquests anys, en els altres es redueix, i viceversa.

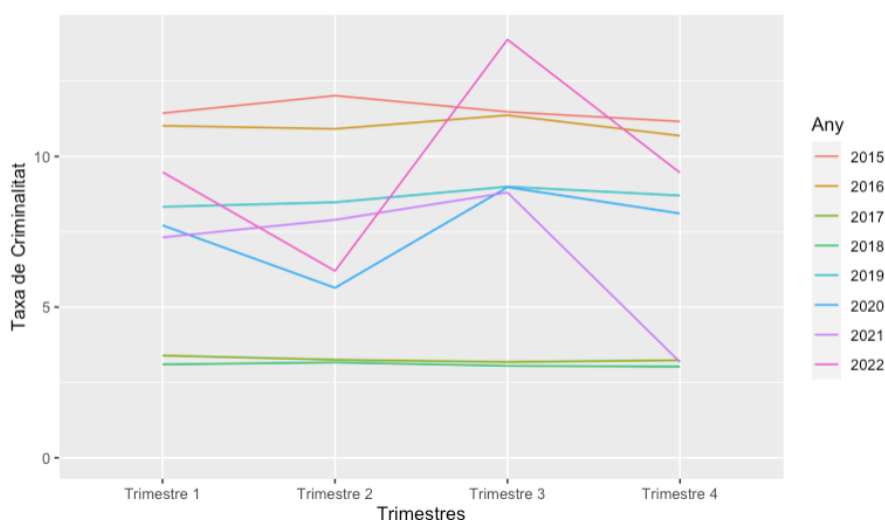
Gràfic 5. Evolució de la taxa de criminalitat trimestral a Navarra. 2015-2022.



Font: Elaboració pròpia a partir de l'R i les dades del Ministeri d'Interior i de l'INE.

D'altra banda, Castella la Manxa si que presenta certs patrons. S'evidencien reduccions generalitzades de la taxa durant el quart trimestre, essent més notables durant els anys 2022 i 2021. De fet és en el quart trimestre del 2018 que es troba el mínim absolut (3,03). Així mateix, el tercer trimestre mostra reduccions generalitzades, essent més intenses en els mateixos anys mencionats anteriorment. Destaca la intensitat de creixement de la taxa de criminalitat durant el tercer trimestre del 2022, convertint-se en un màxim absolut històric (13,88).

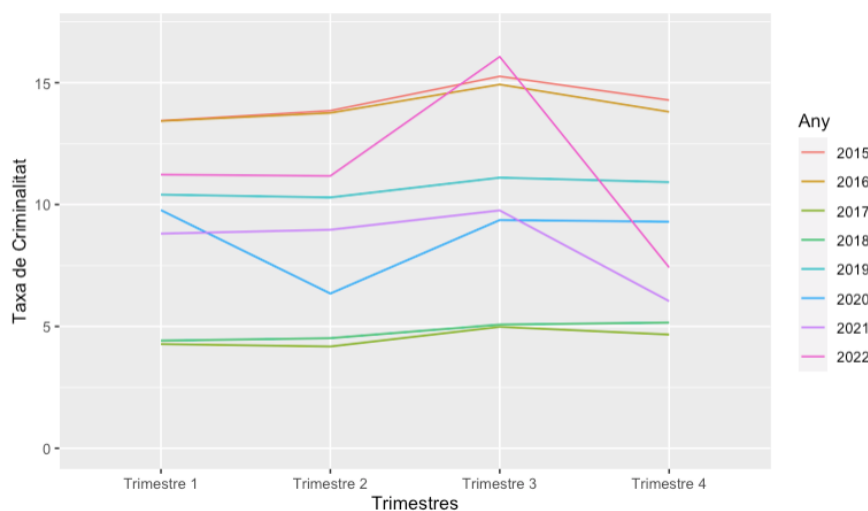
Gràfic 6. Evolució de la taxa de criminalitat trimestral a Castella la Manxa. 2015-2022.



Font: Elaboració pròpia a partir de l'R i les dades del Ministeri d'Interior i de l'INE.

Pel que fa al País Basc, s'observen augments generalitzats del segon trimestre al tercer, mentre que pel quart trimestre s'observen reduccions, excepte per l'any 2020, que es manté constant. De forma similar que a Castella la Manxa, el màxim absolut es troba en el tercer trimestre del 2022 (16,07), mentre que el mínim absolut s'ubica en el primer trimestre del 2017 (4,18).

Gràfic 7. Evolució de la taxa de criminalitat trimestral al País Basc. 2015-2022.

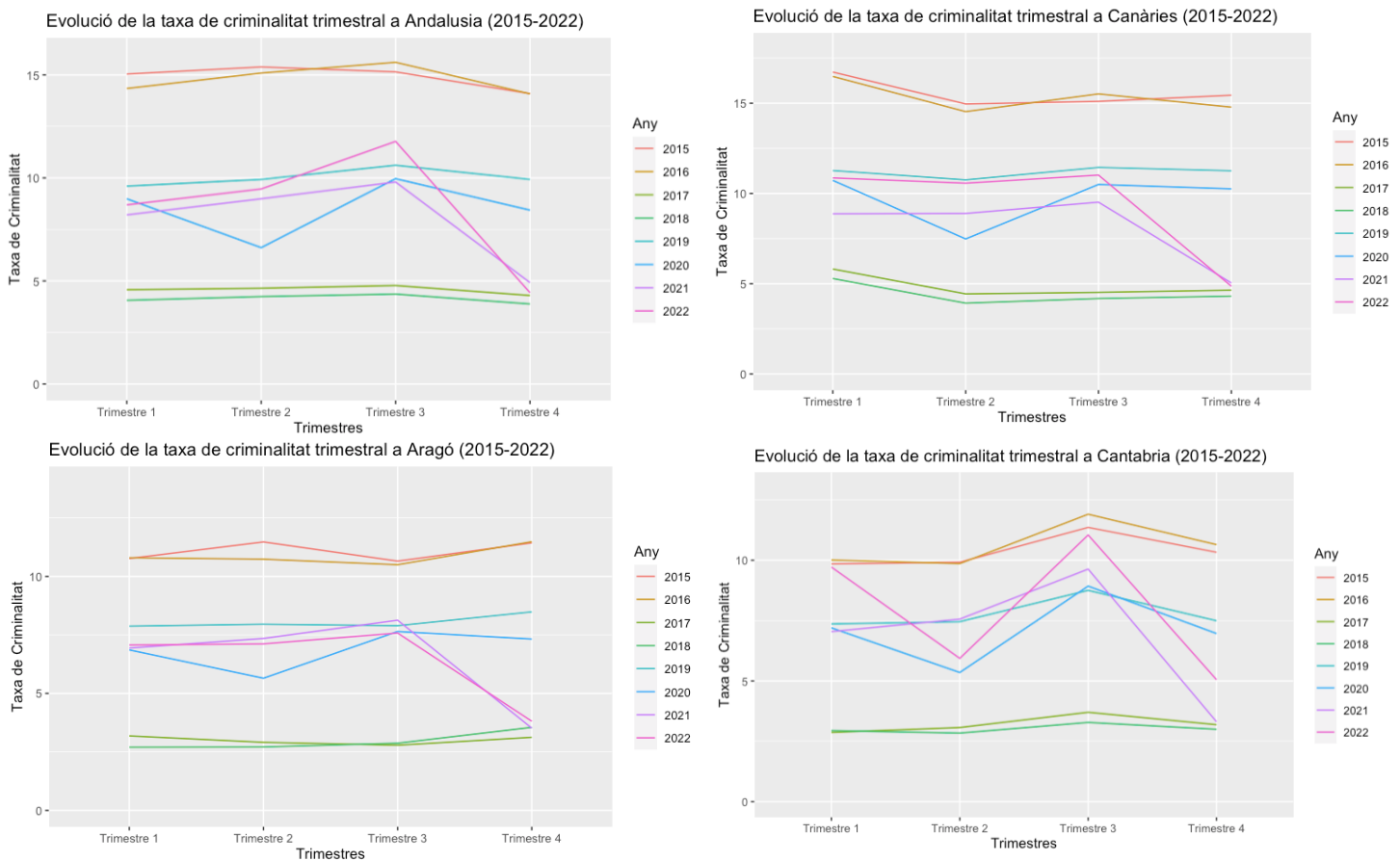


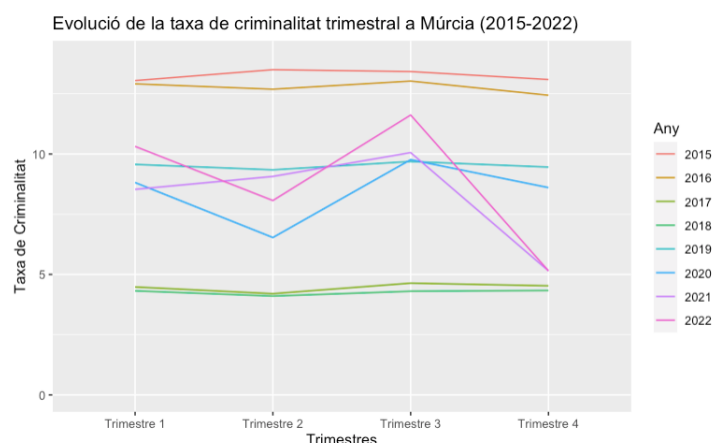
Font: Elaboració pròpia a partir de l'R i les dades del Ministeri d'Interior i de l'INE.

Aquests patrons únics posen de manifest la diversitat en l'evolució de les taxes de criminalitat en aquestes tres regions, alhora que apunten a una certa deterioració en termes de seguretat. Això es deu a la presència de taxes de criminalitat més elevades en els períodes més recents en comparació amb els anys anteriors.

Pel cas d'Andalusia, Canàries (excepte en el 2015), Cantàbria i Múrcia, s'observen reduccions generalitzades en la taxa durant el quart trimestre. Com era d'esperar, en el trimestre anterior, el tercer, la taxa es veu incrementada per totes les CCAA analitzades en el gràfic 8. No obstant això, cal prestar certa atenció a Aragó, que no presenta cap patró trimestral com també succeeix amb les regions de Madrid i Navarra. Pel que fa als valors més extrems, els mínims es troben en el primer trimestre del 2018 pel cas d'Aragó (2,70) i Múrcia (4,11) i al 2017 per Cantàbria (2,84); mentre que en el quart trimestre del 2018 per Andalusia (3,88) i Canàries (3,92). Pel que fa als màxims, hi ha més varietat: per Aragó se situen al quart trimestre del 2016 (11,49); per Cantàbria (11,91) i Andalusia (15,61) al tercer trimestre de 2016; per Canàries al primer trimestre del 2015 (16,73), mentre que per Múrcia és al segon trimestre del mateix any (13,50).

Gràfic 8. Evolució de la taxa de criminalitat trimestral a Andalusia, Canàries, Aragó, Cantàbria i Múrcia. 2015-2022.





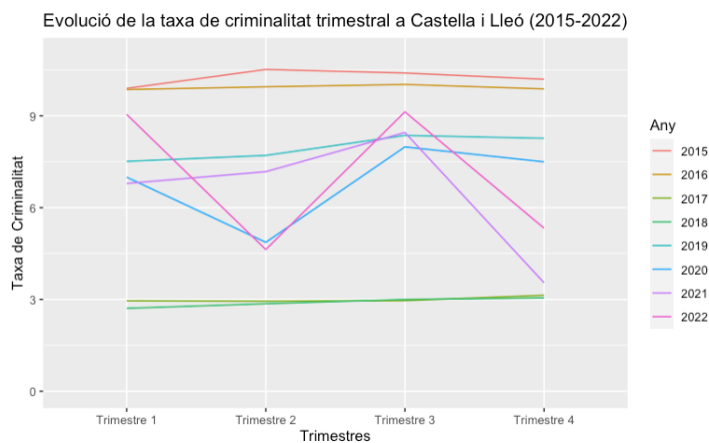
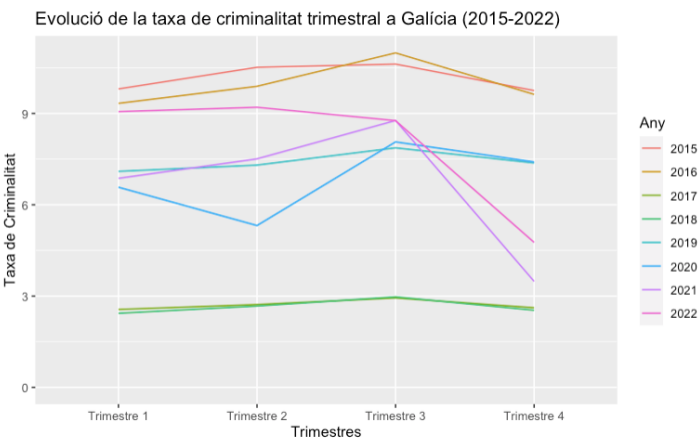
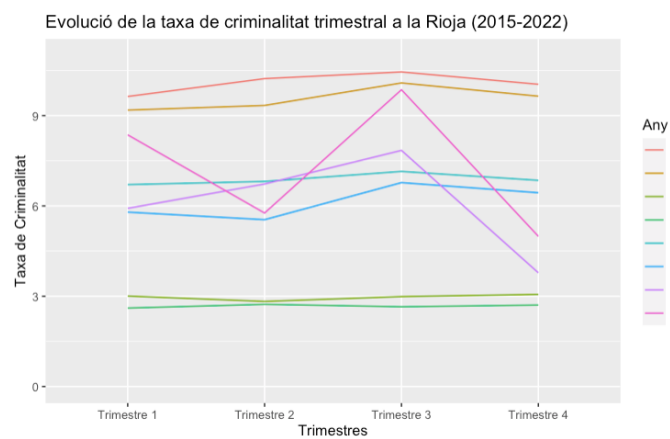
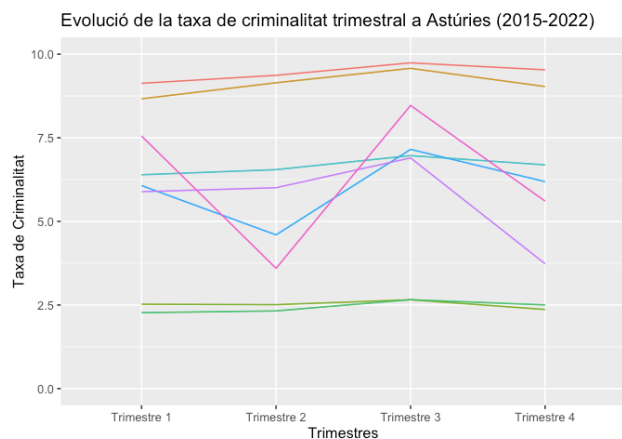
Font: Elaboració pròpia a partir de l'R i les dades del Ministeri d'Interior i de l'INE.

Per finalitzar l'anàlisi trimestral de la taxa de criminalitat per a cada comunitat i ciutat autònoma, es comenten les regions amb les menors taxes.

En general, el quart trimestre presenta una tendència a la baixa en totes les regions, amb algunes excepcions puntuals, com ara Extremadura en determinats anys (2019, 2015, 2016, 2017 i 2018), així com La Rioja i Castella i Lleó (2017 i 2018). D'altra banda, els increments se situen majoritàriament en el tercer trimestre, excepte per a alguns anys en cadascuna de les regions, com és el cas de La Rioja (2018), Extremadura (2015, 2016, 2017 i 2018), Galícia (2015 i 2022) i Castella i Lleó (2015 i 2016).

Els valors més baixos es registren de manera consistent al primer trimestre del 2018. Astúries mostra un valor de 2,27 delictes per 1.000 persones, La Rioja amb 2,61, Extremadura amb 2,17, Galícia amb 2,43 i Castella i Lleó amb 2,71 delictes per 1.000 persones. Pel que fa als màxims, la majoria s'identifiquen en l'any 2015, amb Astúries assolint els 9,74 delictes per 1.000 persones al tercer trimestre, La Rioja amb 10,45, Extremadura amb 9,20 al primer trimestre i Castella i Lleó amb 10,51 al segon trimestre. En canvi, Galícia presenta el seu màxim absolut al tercer trimestre del 2016, amb 10,99 delictes per 1.000 persones.

Gràfic 9. Evolució de la taxa de criminalitat trimestral a Astúries, La Rioja, Extremadura, Galícia i Castella i Lleó. 2015-2022.



Font: Elaboració pròpia a partir de l'R i les dades del Ministeri d'Interior i de l'INE.

4.2 Anàlisi de les tipologies penals

Respecte les tipologies penals de Criminalitat Convencional, s'adjunta una taula d'equivalència dels termes tècnics proporcionats per les dades oficials del Ministeri d'Interior. D'aquesta manera es facilita la comprensió general dels diferents tipus de delictes contemplats a les estadístiques oficials.

Taula 1. Equivalència de nomenclatura tècnica a l'habitual	
Nomenclatura habitual	Nomenclatura tècnica
Assassinat	Homicidis dolorosos i assassinats consumats
Intent d'assassinat	Homicidis dolorosos i assassinats en grau de temptativa
Baralles	Delictes greus i menys greus de lesions i baralla tumultuària
Segrests	Segrests
Violacions	Agressió sexual amb penetració
Agressions sexuals	Resta de delictes contra la llibertat i indemnitat sexual
Robatoris amb violència	Robatoris amb violència i intimidació
Robatoris sense violència	Robatoris amb força en domicilis, establiments i altres instal·lacions
Furts	Furts
Robatori de vehicles	Sostraccions de vehicles
Tràfic de drogues	Tràfic de drogues
Resta de delictes	Resta de criminalitat convencional
Total	Criminalitat convencional total

Cal afegir que la classificació, "resta de criminalitat convencional", es presenta com una categorització general, que inclou tots aquells delictes que no han estat específicament considerats en les altres classificacions.

Per facilitar la comprensió de les taules adjuntes a continuació es subratlla en color verd clar els valors que pertanyen a Espanya, en color taronja els mínims absoluts i en magenta els màxims absoluts. Quan hi ha més de 3 màxims o mínims absoluts no es pinten, ja que es poden trobar fàcilment en la taula. També es facilita la suma total de la taxa del tipus de delicte, sense incloure el valor d'Espanya.

La base de dades oficial presenta diferències significatives en la classificació de les tipologies penals pels anys 2015 i 2016, en comparació amb la classificació presentada a la **Taula 1**. En comptes d'usar la classificació que es mostra anteriorment, usen la següent classificació:

1. Delictes i faltes (EU)
2. Homicidis dolorosos i assassinats consumats (EU)
3. Robatoris amb violència i intimidació (EU)
4. Robatoris amb força en domicilis (EU)
5. Sostraccions de vehicles a motor (EU)
6. Tràfic de drogues (EU)
7. Danys
8. Furts

D'acord amb la metodologia emprada en el recull de les dades, es permet la classificació de "Delictes i faltes (EU)" a "Delictes greus i menys greus de lesions i baralla tumultuària" i els "danys" a la nova classificació de "Resta de criminalitat convencional".

Aquesta divergència en la classificació tindrà repercussions importants en les dades que es mostren a les taules següents, especialment pels segrests, els robatoris sense violència i les agressions sexuals.

• **Assassinats**

Els assassinats són una tipologia penal poc comuns en les regions espanyoles, totes les taxes es situen al voltant del 0. La regió que presenta la major taxa per 1.000 habitants és Ceuta, amb un valor de 0,05 en els anys 2016, 2019 i 2022. Aquests assassinats podrien estar relacionats amb la guerra de narcotraficants que ha estat present en aquesta regió durant varis anys. Aquest fenomen suggereix la influència de factors específics, com el tràfic de drogues, que poden tenir un impacte significatiu en la incidència de crims violents en certes àrees (Carretero, 2022) [6].

Taula 2. Estadístiques d'assassinats. 2015-2022.

	CCAA	2015	2016	2017	2018	2019	2020	2021	2022
1. Homicidis dolorosos i assassinats consumats	Espanya	0,01	0,01	0,01	0,01	0,01	0,01	0,01	0,01
	Catalunya	0,01	0,01	0,01	0,01	0,01	0,01	0,01	0,01
	Illes Balears	0,01	0,01	0,00	0,01	0,01	0,01	0,01	0,00
	Astúries	0,00	0,01	0,00	0,01	0,01	0,00	0,00	0,01
	Aragó	0,01	0,01	0,01	0,01	0,00	0,00	0,00	0,01
	Andalusia	0,01	0,01	0,01	0,01	0,01	0,01	0,01	0,01
	Canàries	0,01	0,01	0,01	0,01	0,01	0,01	0,01	0,01
	Cantàbria	0,01	0,00	0,01	0,00	0,01	0,00	0,01	0,00
	Castella i Lleó	0,00	0,00	0,01	0,00	0,00	0,01	0,01	0,01
	Castella la Manxa	0,01	0,01	0,01	0,00	0,01	0,01	0,00	0,01
	València	0,01	0,01	0,01	0,01	0,01	0,01	0,01	0,01
	Extremadura	0,00	0,01	0,01	0,00	0,00	0,01	0,00	0,00
	Galícia	0,01	0,00	0,00	0,00	0,01	0,00	0,01	0,00
	Madrid	0,00	0,01	0,01	0,01	0,00	0,01	0,01	0,01
	Murcia	0,01	0,01	0,01	0,01	0,01	0,01	0,01	0,01
	Navarra	0,00	0,01	0,02	0,00	0,01	0,00	0,00	0,00
	País Basc	0,01	0,00	0,00	0,00	0,00	0,01	0,00	0,00
	La Rioja	0,00	0,00	0,00	0,00	0,00	0,02	0,01	0,00
	Ceuta	0,01	0,05	0,01	0,01	0,05	0,02	0,02	0,05
	Melilla	0,01	0,01	0,01	0,01	0,01	0,00	0,00	0,02
TOTAL		0,12	0,16	0,13	0,12	0,17	0,13	0,12	0,16

Font: Elaboració pròpia a partir de l'Excel i les dades del Ministeri d'Interior.

- **Intent d'assassinat**

L'intent d'assassinat sembla tenir una presència nul·la o, gairebé nul·la, en els primers anys ja que no es registren dades pels primers anys en les estadístiques oficials. Des del 2016 fins al 2017, es produeix un notable augment en la taxa d'intents d'assassinats que es distribueix arreu de les regions espanyoles, destacant les xifres de les dues ciutats autònomes. Ceuta continua liderant els valors màxims per a tot el període estudiat, llevat de l'any 2018. De fet, registra el màxim històric l'any 2022, amb 0,37 intents d'assassinats per 1.000 habitants. Aquests intents d'assassinats a Ceuta també podrien estar relacionats amb la notable presència de tràfic de drogues a la regió, com es podrà observar a la **taula 11**.

Taula 3. Estadístiques d'intents d'assassinat. 2015-2022.

	CCAA	2015	2016	2017	2018	2019	2020	2021	2022
2. Homicidis dolorosos i assassinats en grau de temptativa	Espanya	0,00	0,00	0,02	0,02	0,02	0,02	0,02	0,03
	Catalunya	0,00	0,00	0,02	0,02	0,02	0,02	0,03	0,03
	Illes Balears	0,00	0,00	0,03	0,03	0,04	0,03	0,03	0,03
	Astúries	0,00	0,00	0,01	0,01	0,01	0,01	0,01	0,02
	Aragó	0,00	0,00	0,01	0,01	0,01	0,03	0,01	0,02
	Andalusia	0,00	0,00	0,02	0,02	0,02	0,03	0,03	0,03
	Canàries	0,00	0,00	0,02	0,02	0,02	0,02	0,02	0,02
	Cantàbria	0,00	0,00	0,01	0,01	0,02	0,01	0,02	0,02
	Castella i Lleó	0,00	0,00	0,01	0,01	0,01	0,01	0,01	0,02
	Castella la Manxa	0,00	0,00	0,02	0,01	0,01	0,02	0,03	0,02
	València	0,00	0,00	0,01	0,02	0,02	0,02	0,02	0,03
	Extremadura	0,00	0,00	0,01	0,01	0,01	0,01	0,02	0,01
	Galícia	0,00	0,00	0,01	0,01	0,01	0,01	0,01	0,02
	Madrid	0,00	0,00	0,01	0,01	0,01	0,01	0,01	0,02
	Murcia	0,00	0,00	0,02	0,02	0,02	0,03	0,03	0,03
	Navarra	0,00	0,03	0,02	0,00	0,02	0,03	0,03	0,03
	Pais Basc	0,00	0,00	0,02	0,02	0,01	0,01	0,02	0,03
	La Rioja	0,00	0,00	0,02	0,01	0,02	0,01	0,02	0,03
	Ceuta	0,00	0,00	0,08	0,01	0,19	0,06	0,12	0,37
	Melilla	0,00	0,00	0,08	0,01	0,09	0,02	0,05	0,06
	TOTAL	0,00	0,03	0,43	0,26	0,56	0,38	0,51	0,82

Font: Elaboració pròpia a partir de l'Excel i les dades del Ministeri d'Interior.

- **Baralles**

Les estadístiques de les baralles revelen una dinàmica peculiar en els dos primers anys, mostrant una taxa notablement elevada en comparació amb la resta dels anys. Aquesta elevada taxa inicial s'atribueix al canvi de classificacions al llarg dels anys, en els anys 2015 i 2016 engloben un nombre més ampli d'infraccions penals que la resta d'anys. Així, malgrat que sembla que la taxa era molt elevada en aquests dos primers anys, en realitat es deu a la inclusió d'altres infraccions que s'han anat segregant de la classificació pels altres anys. Les Illes Balears i Navarra destaquen en aquesta tipologia penal, amb les Balears assolint el seu màxim històric al 2016 amb 61,56 baralles per mil habitants i Navarra amb el seu màxim al 2015 (40,16).

Les baralles acostumen a ser més freqüents en llocs turístics de sol i platja o en zones massificades on la barreja amb substàncies alcohòliques i químiques crea un ambient més tens entre les persones. No sorprèn, per tant, que les zones costaneres del Mediterrani presentin les taxes més elevades en aquesta categoria. No obstant això, en el cas de

Navarra, al 2018 la taxa de baralles arriba a 36,96 per cada mil habitants, mentre que la majoria de les taxes es mantenen prop de zero. D'altra banda, els mínims absoluts es troben l'any 2020 a les regions de l'interior, a Extremadura i a Castella i Lleó.

Taula 4. Estadístiques de les baralles. 2015-2022.

	CCAA	2015	2016	2017	2018	2019	2020	2021	2022
3. Delictes greus i menys greus de lesions i baralla tumultuària	Espanya	43,86	43,23	0,39	0,39	0,42	0,34	0,43	0,55
	Catalunya	55,91	54,17	0,49	0,43	0,44	0,29	0,40	0,52
	Illes Balears	60,37	61,56	0,63	0,71	0,75	0,52	0,65	1,12
	Astúries	25,67	25,11	0,32	0,36	0,39	0,26	0,33	0,45
	Aragó	29,13	29,20	0,30	0,37	0,42	0,30	0,39	0,54
	Andalusia	40,14	39,60	0,35	0,35	0,41	0,36	0,43	0,57
	Canàries	42,63	42,28	0,50	0,50	0,53	0,42	0,50	0,56
	Cantàbria	28,63	29,27	0,42	0,47	0,49	0,40	0,51	0,64
	Castella i Lleó	30,14	29,52	0,27	0,30	0,32	0,24	0,31	0,44
	Castella la Manxa	32,53	31,17	0,28	0,29	0,33	0,29	0,38	0,48
	València	47,13	46,44	0,37	0,40	0,44	0,34	0,48	0,64
	Extremadura	24,93	24,20	0,35	0,32	0,30	0,24	0,29	0,36
	Galícia	28,04	27,42	0,31	0,35	0,35	0,30	0,35	0,44
	Madrid	56,70	57,86	0,33	0,33	0,37	0,34	0,40	0,47
	Murcia	36,38	35,03	0,41	0,42	0,49	0,43	0,57	0,76
	Navarra	40,16	0,61	0,59	36,96	0,61	0,30	0,41	0,61
	País Basc	39,05	37,82	0,48	0,48	0,52	0,40	0,49	0,69
	La Rioja	26,89	25,33	0,41	0,45	0,45	0,38	0,50	0,62
	Ceuta	58,88	54,97	0,68	0,60	0,52	0,90	0,87	0,90
	Melilla	57,21	56,17	0,67	0,59	1,72	1,75	1,38	1,21
	TOTAL	760,52	707,73	8,16	44,67	9,86	8,44	9,65	12,00

Font: Elaboració pròpia a partir de l'Excel i les dades del Ministeri d'Interior.

- **Segrests**

A partir del 2017, s'inicia el registre de dades, tot i que amb taxes molt baixes, destacant les dues ciutats autònomes, Ceuta i Melilla. La tendència general exhibeix variacions ascendents i descendents, i si aquest patró es manté, podríem anticipar un increment en la taxa pel 2023. Cal destacar que el màxim absolut es presenta al 2021, amb 0,036 segrests per mil habitants, un valor que no genera preocupació.

Taula 5. Estadístiques dels segrests. 2015-2022.

	CCAA	2015	2016	2017	2018	2019	2020	2021	2022
4. Segrests	Espanya	-	-	0,001	0,002	0,002	0,002	0,002	0,002
	Catalunya	-	-	0,001	0,001	0,001	0,000	0,001	0,001
	Illes Balears	-	-	0,000	0,001	0,005	0,003	0,006	0,003
	Astúries	-	-	0,000	0,003	0,000	0,001	0,000	0,001
	Aragó	-	-	0,000	0,000	0,002	0,001	0,002	0,001
	Andalusia	-	-	0,002	0,003	0,004	0,003	0,003	0,003
	Canàries	-	-	0,002	0,001	0,000	0,001	0,000	0,000
	Cantàbria	-	-	0,000	0,000	0,000	0,003	0,002	0,000
	Castella i Lleó	-	-	0,001	0,001	0,001	0,001	0,001	0,001
	Castella la Manxa	-	-	0,001	0,002	0,000	0,002	0,001	0,001
	València	-	-	0,001	0,001	0,001	0,003	0,001	0,002
	Extremadura	-	-	0,001	0,000	0,001	0,000	0,000	0,001
	Galícia	-	-	0,001	0,001	0,000	0,000	0,000	0,001
	Madrid	-	-	0,001	0,002	0,002	0,001	0,002	0,002
	Murcia	-	-	0,000	0,002	0,002	0,001	0,002	0,004
	Navarra	-	-	0,005	0,000	0,002	0,006	0,008	0,003
	País Basc	-	-	0,004	0,002	0,003	0,002	0,002	0,002
	La Rioja	-	-	0,000	0,000	0,000	0,000	0,000	0,000
	Ceuta	-	-	0,012	0,000	0,024	0,000	0,036	0,012
	Melilla	-	-	0,012	0,000	0,000	0,000	0,000	0,000
	TOTAL	0,00	0,00	0,04	0,02	0,05	0,03	0,07	0,04

Nota: No es disposa de les dades corresponents al 2015-2016 en la web oficial del Ministeri d'Interior

Font: Elaboració pròpia a partir de l'Excel i les dades del Ministeri d'Interior.

- **Agressions sexuals**

En observar les dades totals, es destaca una tendència a l'alça en el nombre d'agressions sexuals, sense tenir en compte el 2020 que es podria veure afectat per la pandèmia, començant amb 4,67 agressions sexuals per mil habitants al 2017 i arribant a 6,80 al 2022. El màxim absolut es troba al 2022 a la regió de les Illes Balears, la qual lidera els valors màxims d'agressions sexuals per a tots els anys amb registres. A aquesta regió li segueixen, de més lluny, les Canàries i Navarra, amb 0,48 agressions sexuals per mil habitants. Mentre que els mínims absoluts es troben en el 2017 a Astúries, Cantàbria i Extremadura.

L'elevada incidència d'agressions sexuals a les Illes Balears podria estar vinculada a l'ambient festiu i nocturn que caracteritza la zona. La dinàmica intensa de festes i vida nocturna podria crear condicions propícies per a situacions de risc i comportaments inadequats, especialment considerant factors com el consum d'alcohol i les aglomeracions. És crucial prendre en compte aquests elements i implementar estratègies de prevenció i sensibilització en aquests entorns per promoure una atmosfera segura i reduir els casos d'agressions sexuals.

Taula 6. Estadístiques d'agressions sexuals. 2015-2022.

CCAA	2015	2016	2017	2018	2019	2020	2021	2022
Espanya	-	-	0,25	0,29	0,32	0,28	0,32	0,36
Catalunya	-	-	0,29	0,34	0,40	0,30	0,41	0,45
Illes Balears	-	-	0,46	0,52	0,65	0,44	0,61	0,73
Astúries	-	-	0,14	0,20	0,19	0,21	0,22	0,24
Aragó	-	-	0,21	0,26	0,25	0,23	0,29	0,29
Andalusia	-	-	0,25	0,28	0,30	0,26	0,28	0,31
Canàries	-	-	0,30	0,35	0,42	0,40	0,40	0,48
Cantàbria	-	-	0,14	0,18	0,18	0,15	0,20	0,24
Castella i Lleó	-	-	0,15	0,18	0,21	0,17	0,20	0,23
Castella la Manxa	-	-	0,18	0,24	0,26	0,22	0,25	0,27
València	-	-	0,27	0,32	0,35	0,30	0,39	0,43
Extremadura	-	-	0,14	0,20	0,20	0,17	0,18	0,24
Galícia	-	-	0,17	0,20	0,23	0,21	0,23	0,24
Madrid	-	-	0,27	0,31	0,33	0,29	0,32	0,37
Múrcia	-	-	0,27	0,32	0,36	0,34	0,37	0,37
Navarra	-	-	0,42	0,36	0,35	0,36	0,36	0,48
País Basc	-	-	0,23	0,00	0,30	0,27	0,29	0,34
La Rioja	-	-	0,18	0,20	0,29	0,34	0,29	0,32
Ceuta	-	-	0,29	0,59	0,27	0,45	0,54	0,51
Melilla	-	-	0,29	0,58	0,38	0,39	0,44	0,26
TOTAL	0,00	0,00	4,67	5,63	5,92	5,53	6,25	6,80

Nota: No es disposa de les dades corresponents al 2015-2016 en la web oficial del Ministeri d'Interior

Font: Elaboració pròpia a partir de l'Excel i les dades del Ministeri d'Interior.

- **Robatoris amb violència**

Melilla lidera la primera posició amb la major taxa de robatoris amb violència durant els primers anys. Entre el 2017 i el 2018, Catalunya assumeix aquesta posició, seguida de Melilla de nou entre el 2019 i el 2020, i finalment, de Catalunya els últims dos anys. Per tant, els màxims anuals oscil·len entre Melilla i Catalunya. Però, el màxim absolut es registra al 2016 a Melilla (4,39). D'altra banda, el mínim absolut es produeix al 2020 a Extremadura, una de les regions amb menor taxa de criminalitat global.

Taula 7. Estadístiques de robatoris amb violència. 2015-2022.

	CCAA	2015	2016	2017	2018	2019	2020	2021	2022
6. Robatoris amb violència i intimidació	Espanya	1,39	1,36	1,32	1,29	1,39	0,96	1,12	1,33
	Catalunya	2,90	2,70	2,75	2,98	3,47	2,21	2,65	3,23
	Illes Balears	0,97	1,08	1,21	1,09	1,14	0,65	0,76	1,09
	Astúries	0,57	0,51	0,36	0,35	0,35	0,31	0,29	0,39
	Aragó	0,62	0,70	0,61	0,71	0,78	0,59	0,72	0,90
	Andalusia	0,87	0,89	0,92	0,85	0,87	0,62	0,70	0,87
	Canàries	0,73	0,73	0,68	0,70	0,71	0,54	0,61	0,74
	Cantàbria	0,45	0,37	0,37	0,31	0,38	0,31	0,33	0,42
	Castella i Lleó	0,37	0,35	0,30	0,29	0,29	0,23	0,26	0,30
	Castella la Manxa	0,49	0,50	0,51	0,48	0,48	0,37	0,43	0,44
	València	1,24	1,20	1,09	1,03	0,99	0,75	0,91	1,11
	Extremadura	0,28	0,28	0,26	0,23	0,30	0,20	0,21	0,28
	Galícia	0,39	0,40	0,39	0,35	0,35	0,29	0,32	0,36
	Madrid	2,71	2,66	2,51	2,20	2,31	1,59	1,80	1,90
	Murcia	1,01	1,02	0,99	0,98	1,03	0,80	0,91	1,30
	Navarra	0,66	0,67	0,40	0,54	0,51	0,34	0,45	0,63
	Pais Basc	0,87	0,91	0,81	0,81	0,74	0,47	0,64	1,00
	La Rioja	0,43	0,38	0,33	0,25	0,29	0,25	0,36	0,38
	Ceuta	2,06	2,32	1,54	1,76	1,73	1,50	2,33	1,29
	Melilla	3,74	4,39	1,52	1,74	4,34	3,46	2,33	1,43
TOTAL		21,35	22,08	17,54	17,67	21,07	15,47	17,01	18,08

Font: Elaboració pròpia a partir de l'Excel i les dades del Ministeri d'Interior.

- **Robatoris sense violència**

En observar les dades totals, es destaca una tendència a la baixa de la taxa de robatoris en domicilis, establiments i altres instal·lacions. Destaca que Múrcia, Catalunya, les Illes Balears i València lideren en nombre de robatoris sense violència, amb un màxim absolut a Múrcia el 2017 (4,73). Per contra, Ceuta els valors més baixos, assolint un mínim absolut al 2019 (0,68). Cal afegir que Melilla també presenta valors relativament baixos en comparació de la resta de regions.

L'increment dels robatoris en aquestes àrees podria ser atribuïble a factors comuns com el turisme elevat, la concentració urbana, la mobilitat poblacional i la presència d'infraestructures turístiques. Les destinacions turístiques populars solen ser més susceptibles a aquest tipus de delictes, ja que l'afluència de visitants crea oportunitats per als delinqüents, especialment en zones urbanes concorregudes i amb activitats de lleure.

Taula 8. Estadístiques de robatoris sense violència. 2015-2022.

	CCAA	2015	2016	2017	2018	2019	2020	2021	2022
7. Robatoris sense violència	Espanya	-	-	3,21	3,21	3,01	2,18	2,22	2,51
	Catalunya	-	-	4,16	4,57	4,43	2,86	2,95	3,60
	Illes Balears	-	-	3,84	3,61	3,11	2,20	2,27	2,46
	Astúries	-	-	1,68	1,45	1,53	1,10	1,19	1,19
	Aragó	-	-	1,95	1,91	1,78	1,42	1,53	1,81
	Andalusia	-	-	2,95	2,75	2,67	2,08	1,93	2,20
	Canàries	-	-	2,54	2,36	2,00	1,60	1,58	1,68
	Cantàbria	-	-	2,82	2,71	2,80	2,36	2,79	3,10
	Castella i Lleó	-	-	2,14	2,22	1,99	1,40	1,59	1,84
	Castella la Manxa	-	-	3,82	3,65	3,21	2,32	2,16	2,34
	València	-	-	4,36	4,40	3,96	3,01	3,15	3,53
	Extremadura	-	-	2,03	1,88	1,55	1,48	1,38	1,74
	Galícia	-	-	2,01	1,98	1,96	1,50	1,39	1,53
	Madrid	-	-	2,97	2,94	2,74	1,81	1,83	2,02
	Murcia	-	-	4,73	4,57	4,14	3,11	3,21	3,49
	Navarra	-	-	2,35	2,37	2,31	1,66	1,91	2,20
	Pais Basc	-	-	3,31	3,64	3,70	2,51	2,97	3,43
	La Rioja	-	-	2,29	1,90	1,74	1,42	1,55	2,22
	Ceuta	-	-	0,98	1,03	0,68	0,82	0,71	0,54
	Melilla	-	-	0,96	1,02	1,13	0,78	0,81	1,01
TOTAL		0,00	0,00	51,88	50,95	47,43	35,42	36,91	41,93

Nota: No es disposa de les dades corresponents al 2015-2016 en la web oficial del Ministeri d'Interior

Font: Elaboració pròpia a partir de l'Excel i les dades del Ministeri d'Interior.

- Furts**

Entenent els furts com delictes en què una persona pren o extreu propietat d'una altra sense usar la força ni la violència directa, es revela com el delicte més comú a Espanya. Aquesta tendència és especialment evident a les regions amb majors taxes de criminalitat, com Catalunya, les Illes Balears i Madrid. De fet, el màxim absolut es troba a les Illes Balears amb 26,15 furts per mil habitants l'any 2017. Les dades indiquen que les regions amb més casos de furts són aquelles amb més incidència global de delictes. En contrast, les regions amb menor taxa de criminalitat, com Extremadura i Astúries, presenten valors més baixos de furts, amb un mínim absolut de 3,96 furts per mil habitants a Extremadura l'any 2020.

Taula 9. Estadístiques de furts. 2015-2022.

	CCAA	2015	2016	2017	2018	2019	2020	2021	2022
8. Furts	Espanya	15,41	15,30	15,27	15,03	14,80	8,89	10,39	13,37
	Catalunya	22,28	21,86	23,14	25,43	25,17	14,31	15,32	20,88
	Illes Balears	25,05	25,23	26,15	24,36	23,13	10,66	14,01	20,79
	Astúries	7,87	7,33	7,02	7,01	6,63	4,07	4,65	6,78
	Aragó	9,22	9,13	8,41	8,12	7,99	5,09	6,73	8,05
	Andalusia	13,19	13,06	12,38	11,13	10,85	6,80	8,13	10,48
	Canàries	14,40	14,75	14,29	12,82	12,13	7,68	7,89	10,85
	Cantàbria	8,71	8,79	8,59	8,02	8,38	5,55	6,54	8,34
	Castella i Lleó	9,90	9,19	8,69	8,18	8,00	4,77	5,78	7,32
	Castella la Manxa	8,29	7,72	7,48	7,13	6,80	4,71	5,35	7,02
	València	15,83	15,57	14,96	14,12	13,61	8,67	11,41	13,99
	Extremadura	7,40	6,86	6,71	6,16	6,15	3,96	4,55	6,06
	Galícia	8,06	7,86	7,50	7,34	6,98	4,58	5,70	7,27
	Madrid	22,46	23,28	24,02	23,20	23,14	13,60	16,11	19,68
	Murcia	11,38	10,75	10,38	9,97	10,13	6,56	7,85	10,10
	Navarra	12,90	11,82	12,04	12,11	13,05	6,76	8,35	12,21
	Pais Basc	12,88	13,22	12,70	13,36	13,70	8,07	10,05	13,24
	La Rioja	9,07	8,83	8,13	7,42	7,62	4,61	5,59	8,58
	Ceuta	12,66	11,30	9,53	10,15	9,66	5,62	5,20	5,19
	Melilla	15,52	15,26	9,41	10,00	13,90	8,93	6,68	8,59
TOTAL		247,07	241,80	231,53	226,04	227,04	135,00	155,91	205,43

Font: Elaboració pròpia a partir de l'Excel i les dades del Ministeri d'Interior.

- **Robatori de vehicles**

Ceuta lidera la classificació com la regió amb més robatoris de vehicles durant el període analitzat, registrant un màxim històric de 4,65 robatoris de vehicles per mil habitants al 2016. No obstant això, a partir d'aquest any, la taxa disminueix de manera considerable i s'aproxima als nivells d'altres regions, com les Illes Balears, que arriba, fins i tot, a superar a Ceuta al 2022 (1,33). A diferència d'aquestes dues, Aragó, Navarra i La Rioja presenten taxes notablement més baixes, sent a La Rioja on s'assoleix el mínim absolut al 2020 amb només 0,10 robatoris de vehicles per mil habitants.

Les elevades taxes de robatoris de vehicles a Ceuta podrien atribuir-se a la seva ubicació geogràfica com a ciutat autònoma propera del nord d'Àfrica, facilitant la possibilitat de desplaçament cap a un altre continent a través de les fronteres. Alhora, la densitat de població, la presència d'infraestructures portuàries i l'accessibilitat també podrien influir en aquest fenomen. Pel que fa a les Illes Balears, la significativa afluència turística podria contribuir a l'increment dels robatoris de vehicles, ja que les zones turístiques sovint atrauen delinqüents.

Taula 10. Estadístiques de robatori de vehicles. 2015-2022.

	CCAA	2015	2016	2017	2018	2019	2020	2021	2022
9. Sostraccions de vehicles	Espanya	0,93	0,94	0,91	0,77	0,74	0,55	0,56	0,64
	Catalunya	1,16	1,07	1,18	1,22	1,27	0,81	0,80	0,96
	Illes Balears	1,60	1,75	1,54	1,28	1,24	0,89	0,91	1,33
	Astúries	0,43	0,36	0,38	0,21	0,22	0,29	0,26	0,25
	Aragó	0,28	0,30	0,30	0,22	0,18	0,16	0,19	0,21
	Andalusia	1,07	1,06	1,04	0,83	0,77	0,58	0,57	0,66
	Canàries	0,80	0,77	0,85	0,65	0,60	0,47	0,45	0,57
	Cantàbria	0,43	0,42	0,36	0,21	0,23	0,18	0,24	0,15
	Castella i Lleó	0,32	0,37	0,27	0,24	0,24	0,18	0,20	0,22
	Castella la Manxa	0,62	0,74	0,54	0,33	0,30	0,25	0,25	0,32
	València	0,87	0,89	0,79	0,67	0,64	0,49	0,49	0,63
	Extremadura	0,35	0,38	0,31	0,15	0,13	0,16	0,14	0,16
	Galícia	0,36	0,38	0,29	0,22	0,22	0,17	0,15	0,20
	Madrid	1,57	1,63	1,65	1,32	1,21	0,94	1,03	1,10
	Murcia	0,91	0,93	0,88	0,65	0,61	0,49	0,44	0,51
	Navarra	0,34	0,26	0,27	0,29	0,28	0,20	0,18	0,20
	País Basc	0,40	0,45	0,41	0,42	0,40	0,29	0,31	0,33
	La Rioja	0,42	0,31	0,28	0,25	0,22	0,10	0,18	0,25
	Ceuta	4,15	4,65	2,54	2,91	1,98	1,00	1,23	1,29
	Melilla	1,40	1,22	2,51	2,87	0,88	0,49	0,41	0,94
	TOTAL	17,49	17,95	16,40	14,94	11,63	8,14	8,44	10,27

Font: Elaboració pròpia a partir de l'Excel i les dades del Ministeri d'Interior.

- **Tràfic de drogues**

Les dues ciutats autònomes lideren les regions amb més tràfic de drogues, destacant Ceuta amb un màxim històric de 4,60 detencions per tràfic de drogues per cada mil habitants al 2015. Les zones amb menys presència de tràfic de drogues són Castella i Lleó, Galícia i Astúries, amb un mínim històric de 0,10 al 2016. Sorprenen especialment les estadístiques a Galícia, ja que en les dècades dels 80, aquesta regió era la principal porta d'entrada a Europa de les drogues procedents de Colòmbia i encara, avui en dia, popularment s'identifica com una àrea amb molta presència de narcotraficants.

Taula 11. Estadístiques de tràfic de drogues. 2015-2022.

	CCAA	2015	2016	2017	2018	2019	2020	2021	2022
10. Tràfic de drogues	Espanya	0,26	0,27	0,28	0,30	0,35	0,36	0,39	0,40
	Catalunya	0,26	0,29	0,30	0,32	0,34	0,38	0,50	0,51
	Illes Balears	0,38	0,45	0,43	0,46	0,57	0,36	0,33	0,53
	Astúries	0,15	0,10	0,13	0,14	0,20	0,17	0,16	0,17
	Aragó	0,22	0,19	0,21	0,26	0,27	0,24	0,31	0,31
	Andalusia	0,37	0,37	0,41	0,44	0,52	0,55	0,57	0,54
	Canàries	0,27	0,28	0,28	0,29	0,28	0,27	0,24	0,34
	Cantàbria	0,20	0,16	0,11	0,15	0,15	0,16	0,16	0,14
	Castella i Lleó	0,13	0,14	0,14	0,17	0,19	0,21	0,21	0,20
	Castella la Manxa	0,17	0,18	0,21	0,22	0,23	0,27	0,28	0,27
	València	0,30	0,33	0,32	0,35	0,40	0,37	0,39	0,40
	Extremadura	0,18	0,18	0,19	0,22	0,30	0,27	0,31	0,25
	Galícia	0,13	0,13	0,14	0,16	0,19	0,20	0,17	0,19
	Madrid	0,22	0,23	0,25	0,30	0,39	0,35	0,37	0,44
	Murcia	0,17	0,18	0,20	0,19	0,24	0,26	0,33	0,34
	Navarra	0,31	0,21	0,22	0,23	0,24	0,22	0,22	0,22
	Pais Basc	0,16	0,15	0,15	0,17	0,19	0,25	0,24	0,28
	La Rioja	0,19	0,20	0,25	0,23	0,33	0,23	0,23	0,31
	Ceuta	4,60	3,04	2,67	1,95	3,50	2,61	2,45	1,38
	Melilla	1,31	0,95	2,64	1,92	0,88	2,65	2,67	0,79
TOTAL		9,72	7,77	9,24	8,16	9,41	10,04	10,15	7,62

Font: Elaboració pròpia a partir de l'Excel i les dades del Ministeri d'Interior.

- Resta de delictes**

En aquesta classificació s'inclouen totes les tipologies penals que no pertanyen a la resta de classificacions, constituint, per tant, una categoria àmplia que engloba una gran varietat de delictes, denominada comunament "el calaix desastre".

En la majoria de les regions analitzades, s'observa un dràstic increment de les taxes al 2017, fenomen que podria ser atribuït al canvi de classificació prèviament assenyalat. Com era d'esperar, aquelles regions amb una taxa de criminalitat més elevada exhibeixen els valors més pronunciats en la resta de delictes considerats en aquesta categoria. Destaquen, en aquest sentit, Melilla, Madrid, València, Catalunya, les Illes Balears, Ceuta i Navarra. Per contra, les regions amb valors més moderats són Astúries, Cantàbria, Extremadura i La Rioja. El màxim absolut es registra el 2019 a Melilla, amb una taxa de 41,21, mentre que el mínim correspon al 2016 i es localitza a Astúries.

Taula 12. Estadístiques de la resta de delictes. 2015-2022.

	CCAA	2015	2016	2017	2018	2019	2020	2021	2022
11. Resta de criminalitat convencional	Espanya	4,64	4,62	22,19	24,10	25,40	23,84	19,34	21,38
	Catalunya	5,24	5,15	23,92	27,09	28,58	24,44	20,89	22,77
	Illes Balears	4,52	4,58	30,36	32,54	36,93	35,50	25,36	29,14
	Astúries	2,96	2,90	14,75	15,99	17,03	17,51	12,09	13,95
	Aragó	4,06	4,00	16,94	18,89	20,59	19,36	14,95	16,83
	Andalusia	4,00	4,11	21,56	23,09	23,90	22,79	19,50	22,26
	Canàries	3,41	3,45	23,87	24,87	27,71	27,32	20,69	22,31
	Cantàbria	2,96	3,36	17,23	18,36	18,47	19,34	16,77	19,22
	Castella i Lleó	4,32	4,36	17,67	19,42	20,57	20,08	15,49	17,08
	Castella la Manxa	3,84	3,60	18,45	21,25	22,95	22,02	18,07	20,62
	València	5,61	5,49	23,76	25,34	26,06	24,60	21,47	24,16
	Extremadura	2,92	2,92	14,45	15,55	16,89	17,09	13,43	14,20
	Galícia	3,63	3,59	16,56	18,24	19,34	20,03	16,24	17,27
	Madrid	6,81	6,81	26,58	28,66	30,70	27,08	21,50	24,13
	Murcia	3,18	3,22	17,95	20,00	21,19	21,76	19,19	21,44
	Navarra	5,06	22,40	26,09	4,88	27,20	25,80	21,86	25,45
	Pais Basc	3,47	3,42	20,02	21,82	23,23	22,46	15,68	17,27
	La Rioja	3,29	3,18	13,09	14,66	16,66	17,20	13,18	16,48
	Ceuta	3,88	3,53	34,65	35,87	37,37	30,55	27,19	25,51
	Melilla	4,07	4,36	34,18	35,35	41,21	33,12	27,92	31,67
TOTAL		77,22	94,42	412,10	421,89	476,57	448,06	361,47	401,77

Font: Elaboració pròpia a partir de l'Excel i les dades del Ministeri d'Interior.

4.3 Conclusions de l'anàlisi descriptiva

De l'anàlisi temporal i geogràfica, destaquen Ceuta, Madrid, Catalunya i les Illes Balears com les comunitats autònomes i ciutat autònoma amb major criminalitat. Per contra, Astúries, Extremadura i La Rioja presenten les menors taxes de criminalitat.

Les diferències regionals revelen altes taxes en destins turístics amb major densitat poblacional i baixes taxes associades a regions amb menor densitat poblacional i activitat econòmica. Aquestes troballes destaquen la complexitat del fenomen criminal, posant de relleu la necessitat d'un anàlisi més detallat per a comprendre les seves dinàmiques.

4.3.1 Conclusions anàlisi descriptiva trimestral

De l'anàlisi de l'evolució trimestral, s'observen patrons distintius que agrupen diverses comunitats autònomes. Per una banda, Catalunya i les Illes Balears comparteixen similituds amb augments al tercer trimestre i reduccions al quart, possiblement relacionades amb el turisme. D'altra banda, Madrid i Ceuta mostren comportaments atípics, amb fluctuacions destacades el 2020 i el 2022, mentre que València i Melilla comparteixen tendències trimestrals d'augment i reducció. Navarra, Aragó i Madrid presenten una falta de patrons clars, assenyalant la complexitat dels factors que afecten la criminalitat. Seguidament, el País Basc, Andalusia, Canàries, Cantàbria i Múrcia comparteixen reduccions al quart trimestre, indicant possibles influències estacionals. Finalment, Astúries, La Rioja, Extremadura, Galícia i Castella i Lleó, les regions amb taxes més baixes, revelen una tendència a la baixa al quart trimestre i augments al tercer, reflectint possibles factors socioeconòmics comuns.

Aquestes associacions suggereixen que el turisme, les característiques urbanes, les variables socioeconòmiques i l'estacionalitat són factors clau que influeixen en els patrons de criminalitat en aquestes regions.

4.4 Conclusions de l'anàlisi de tipologies penals

Les tipologies penals més destacades a les regions espanyoles, en ordre descendent, són la "resta de criminalitat convencional", seguida dels furts, els robatoris amb força i a domicilis, i les baralles. Aquesta última classificació revela un canvi significatiu des del 2018, que es pot atribuir a una nova classificació creada pel Ministeri de l'Interior. Aquest canvi ha desviat moltes baralles que anteriorment es consideraven en aquesta categoria cap al grup de "resta de criminalitat convencional", el qual ha experimentat un augment notable des de llavors.

Les dades revelen que Ceuta ocupa la primera posició en diverses tipologies penals, incloent-hi assassinats, intents d'assassinats, robatori de vehicles i tràfic de drogues. Melilla destaca en robatoris amb violència i tràfic de drogues. Una altra regió que lidera la primera posició són les Illes Balears, destacada en agressions sexuals, furts, robatoris sense violència (juntament amb Múrcia, Catalunya i València) i baralles (juntament amb Navarra). Finalment, les regions de Melilla, Madrid, València i Catalunya encapçalen les estadístiques en la categoria de "resta de delictes".

Aquest panorama destaca la diversitat de patrons delictius i la seva distribució geogràfica. Les regions costaneres, com les Illes Balears i Ceuta, mostren una major incidència de delictes vinculats al turisme i a la vida nocturna. D'altra banda, les regions urbanes, com Catalunya i Madrid, destaquen en una àmplia gamma de tipologies penals, indicant la complexitat dels factors que influeixen en els nivells de criminalitat.

V. MODELS ESTADÍSTICS

En aquest capítol es tractarà d'explicar els models estadística que es faran servir per analitzar quins factors influeixen en el nombre de crims convencionals. Per a realitzar la explicació d'aquests models, es recorre a les presentacions didàctiques de Comptatges de l'assignatura Models Lineals Generalitzats, J. Cortès (2023).

Els models estadístics pretenen aproximar-se el màxim possible a la realitat, per aquest motiu són eines usades per observar l'associació entre les variables que es volen estudiar (variables explicatives/independents), i el fenomen d'interès, la variable resposta (o d'interès/dependent).

Els models lineals generalitzats estan constituïts per un component aleatori, que es tracta del vector de valors que pren la variable resposta (Y), un component sistemàtic (η) i una funció d'enllaç ($g(\mu)$) que es relaciona amb el predictor lineal amb el valor esperat (μ) de la resposta condicionat al valor que prenen les variables explicatives.

$$g(\mu_i) = \eta_i = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_m X_m$$

En aquest projecte s'usaran tres models estadístics per a veure quina factors expliquen el número de criminalitat convencional, per una banda, i el número de delictes que afecten directament a la persona, per l'altra.

5.1 Models Log-lineals

5.1.1 Model Poisson

Els models log-lineals són un tipus de model estadístic usat per analitzar dades de comptatge, on la variable d'interès és el nombre d'ocurrències d'un esdeveniment en una unitat d'observació, en el nostre cas el nombre de crims. Es basa en la distribució de Poisson i usa la transformació logarítmica per a modelar la relació entre les variables predictores i la d'interès.

$$\log(\mu_i) = \theta = \eta_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \dots + \beta_m X_{im}$$

Els models log-lineals posseeixen dues característiques:

- 1- La dependència de l'esperança de la resposta (μ_i) condicionada a un vector de covariables i/o factors (x_i) és multiplicativa i generalment es formula en forma logarítmica.

$$\log(y_i) = \log(t_i) + x_i^T \beta \quad i = 1, \dots, n$$

- 2- S'assumeix que la variància condicional de la resposta és proporcional a la seva esperança i constant en tot el conjunt de dades:

$$V(Y_i|X_i) = \phi \cdot \mu_i$$

On els valors de ϕ ens indicarà quin tipus de dispersió hi haurà al model.

Infradispersió	Equidispersió	Sobredispersió
$\phi < 1$	$\phi = 1$	$\phi > 1$

Sovint, interessa saber la taxa de la variable d'interès, per a aconseguir-la només cal afegir el terme d'offset. Es tracta d'una variable que s'introdueix en el model, habitualment, en forma logarítmica del temps de seguiment (t) o de la mida del grup (n) per a poder obtenir la taxa:

$$\log(y_i) = \log(t_i) + x_i^T \beta \rightarrow \log\left(\frac{y_i}{t_i}\right) = \log(\mu_i) = x_i^T \beta$$

L'enllaç és la funció que connecta la mitjana de la distribució, en aquest cas, μ_i amb els predictors lineals (η_i) del model. Pel model de regressió de Poisson el logaritme és l'enllaç canònic (θ). Això vol dir que es modela el logaritme natural de la taxa mitja ($\theta = \log(\mu_i)$) en funció de les variables predictores. Permetent, d'aquesta manera que la relació entre les variables predictores i la taxa mitjana sigui lineal.

Les estimacions dels paràmetres es fan a partir de la funció de versemblança (L), fent que les estimacions siguin asimptòticament consistents i normals. El que significa que, per a mostres prou grans, les seves distribucions s'aproximen a distribucions normals.

$$L(\alpha, \beta_1, \dots, \beta_m | Y, X) = \prod_{i=1}^n \frac{e^{-\mu_i} \mu_i^{y_i}}{y_i!}$$

On y_i és el nombre observat d'esdeveniments per a la i-èsima observació.

En un model logístic, la probabilitat (p) de que un esdeveniment (per exemple, la ocurrència d'un delict menor / greu) es relaciona amb les variables independents a través de la funció logística:

$$p = \frac{1}{1 + e^{-(\beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \dots + \beta_m X_{im})}}$$

Pel que fa a la interpretació dels paràmetres del model, és fa en termes d'odd ratio (OR). Els OR indicaran com la probabilitat d'ocurrència d'un delict menor (o greu, depenent de la variable resposta) canvia en relació amb les variables independents en el model.

El terme constant (intercepta, β_0) representa la taxa mitjana global estimada del esdeveniment (si incloem l'offset, com és el nostre cas) quan les variables explicatives són nul·les. En presència de variables explicatives, la exponencial de la constant (e^{β_0}) indica la taxa mitja estimada del grup de referència. Per a variables categòriques, la exponencial dels coeficients s'interpreta com el canvi multiplicatiu en la taxa mitjana en comparació de la categoria de referència, essent major a 1 si es produeix un increment i menor a 1 si es tracta d'una disminució.

Hi ha diferents maneres de comprovar la bondat d'ajust d'un model log-lineal i la seva validació, tant numèriques com gràfiques. Una d'elles, numèrica, és l'estadístic de Pearson (χ^2) que s'usa per a avaluar la discrepància entre les freqüències observades (O_i) i les freqüències esperades (E_i) sota el model ajustat.

$$\chi^2 = \sum_{i=1}^n \frac{(O_i - E_i)^2}{E_i} \sim \chi_{J-p-1}^2$$

Aquest estadístic segueix aproximadament una distribució chi-quadrat (χ_{J-p-1}^2) sota la hipòtesis nul·la de que el model ajustat és correcte, on J és el nombre d'interval·ls per calcular l'estadístic i p el nombre de paràmetres del model. Quan el p-valor associat a l'estadístic és menor que un nivell de significança escollit (normalment 0.05), es rebutja la hipòtesi nul·la, indicant que el model no s'ajusta adequadament a les dades.

Una altra mesura numèrica és l'estadístic de la deviància (D) que es tracta d'una mesura de discrepància entre el model ajustat i el nul.

$$D = 2 \sum_{i=1}^n O_i \log \left(\frac{O_i}{E_i} \right) \sim \chi_{J-p-1}^2$$

Aquest estadístic també segueix una distribució chi-quadrat (χ_{J-p-1}^2) sota la hipòtesis nul·la de que el model ajustat és correcte.

Les dues proves es basen en comparar el nostre model amb el model saturat com també són asimptòticament equivalents i haurien de coincidir raonablement per a mides mostrals grans.

La validació gràfica amb els residus s'usa per a avaluar la qualitat de l'ajust d'un model estadístic, com també es proporciona informació vital sobre la validesa de les suposicions del model i la presència de patrons sistemàtics en les dades.

- **Gràfic de residus vs. Valors ajustats:** mostra els residus en l'eix vertical i els valors ajustats pel model en l'eix horitzontal. Si el model és adequat, no hauria d'haver patrons evidents en la distribució dels residus a mesura que els valors ajustats augmenten, com també que els residus han d'estar distribuïts aleatòriament al voltant del 0. Al R es fa a partir del paquet *car* i la funció *ResidualPlot*.
- **Gràfic Quantil-Quantil (Q-Q Plot) de residus:** compara els quantils teòrics d'una distribució concreta (com la normal) amb els quantils observats dels residus. Si els residus segueixen la distribució concreta, els punts haurien d'aproximar-se a una línia diagonal. Desviacions significatives poden indicar problemes amb la normalitat dels residus. Al R es fa a partir del paquet *base*.
- **Gràfic de residus vs. Variables explicatives del model:** permeten detectar patrons no lineals o violacions de l'homoscedasticitat. Si hi ha patrons sistemàtics, podria ser necessària la transformació de les variables (afectant a la interpretació dels coeficients). Al R es fa a partir del paquet *car* i la funció *ResidualPlots*.
- **Gràfic de residus vs. Variables explicatives no incloses al model:** útil per a identificar variables importants que no s'han inclòs al model. Quan s'observen patrons sistemàtics acostuma a ser perquè es necessiten incloure variables addicionals al model. Al R es fa a partir del paquet *base*.
- **Residus vs. Eix temporal:** si s'observen patrons en la gràfica de residus respecte el temps, podria indicar falta d'independència temporal. Al R es fa a partir del paquet *stats*.

S'implementa la funció *glm* del paquet *stats* (R Core Team, 2020) [30], declarant la família com Poisson i l'enllaç el logaritme. A continuació es mostra com es faria en l'R, on fórmula seria el model que es desitja estudiar, amb la variable resposta i les variables explicatives corresponents.

Figura 1. Funció *glm* del paquet *stats* en R

```
glm(formula,dades,family=poisson)
```

El model de Poisson assumeix que la mitjana i la variància de la variable resposta són iguals, és a dir que el paràmetre de dispersió ($\phi = 1$). Però, a la pràctica aquesta suposició no acostuma a complir-se, és per això que sempre cal fer un test de sobredispersió. Per solucionar aquesta problemàtica existeixen models alternatius, depenent de la relació que hi hagi entre la variància i l'esperança s'usarà el model Quasi-Poisson o el Binomial negatiu.

5.1.2 Model Quasi-Poisson

Aquest model relaxa la restricció d'igualtat entre la mitjana i la variància, permetent que la **relació** sigui **lineal**, però no necessàriament constant. Dit d'una altra manera, quan la sortida del test de sobredispersió només surt significativa només per a *trafo* = 1.

Prèviament a l'estimació, es realitza una estimació preliminar del paràmetre de dispersió: $\hat{\phi} = \frac{\chi_p^2}{n-p}$ on χ_p^2 és l'estadístic de *Pearson*. Aquesta estimació podria aproximar-se raonablement mitjançant la següent expressió: $\hat{\phi} = 1 + \alpha$ on α és l'estimació procedent quan *trafo* = 1, per tant, quan $V(Y|X) = (1 + \alpha) \cdot \mu_i$.

Els coeficients del model s'estimen usant la distribució de Poisson, per tant, les estimacions puntuals són les mateixes que en el model de Poisson. Però els errors estàndards s'inflen de forma constant pel paràmetre de dispersió estimat ($\sqrt{\phi}$).

Atès que no és un model probabilístic, no és possible calcular la versemblança ni els estadístics com la Deviància, l'AIC o el BIC. No obstant això, s'empra un test de la Deviància modificada per a comparar models Quasi-Poisson niats, expressat com:

$$\frac{D_2 - D_1}{\phi(p_1 - p_2)} \sim F_{p_1 - p_2, n - p_1}$$

On ϕ és el paràmetre de sobredispersió estimat amb el model amb més paràmetres.

Figura 2. Funció *glm* del paquet *stats* en R pel Model Quasi-Poisson

```
glm(formula, family=quasipoisson, data = dades)
```

5.1.3 Model Binomial Negatiu

La distribució binomial negativa és una distribució de probabilitat discreta que representa la variabilitat en el nombre d'ocurrències d'un esdeveniment fins que s'aconsegueix un nombre predeterminat d'èxits. Es considera aquest model quan la **relació** entre la **mitjana** i la **variància** és **quadràtica**, $V(Y|X) = E(Y|X) + \alpha_2 \cdot E(Y|X)^2$, dit d'una altra manera, quan al test de sobredispersió surt significatiu *trafo* = 2, independentment de si *trafo* = 1 és o no significatiu.

Primerament cal estimar el paràmetre de dispersió, escollint aquell que maximitza la versemblança. A la pràctica es realitza usant la funció *glm.nb* del paquet de *MASS* en R. Això s'aconsegueix ajustant un model inicial amb la funció *glm.nb* i extraient el valor estimat de ϕ

(*model1\$theta*). Aquesta estimació s'usarà per a formular el model complert usant la funció *glm* amb el paràmetre *family* establert com *negative.binomial(Theta)*.

Figura 3. Funció *glm* i *glm.nb* del paquet *stats* en R pel Model Binomial Negatiu

```
mod.BN <- glm.nb(formula, data=dades)
mod.BN1 <- glm(formula, data=dades, family = negative.binomial(mod.BN$theta))
```

Depenent de la naturalesa de les dades es poden considerar altres models per a millorar l'adequació del model a la realitat dels comptatges observats, per exemple, si hi ha més 0 dels esperats en la variable resposta, es podria considerar un model zero-inflat, o si per altra banda, no hi ha 0 en la variable resposta i n'hauríem d'esperar uns quants, es podria considerar aplicar el model zero-truncat.

Respecte la interpretació dels models, és molt semblant a la dels anteriors models. La constant del model representa el logaritme del nombre esperat d'èxits (ocurrències de l'esdeveniment) quan totes les variables explicatives són 0 (en condicions basi). Per a les variables explicatives, els coeficients associats indiquen el canvi en el logaritme del nombre esperat d'èxits per a un augment unitari en la variable, mantenint constants les altres variables. De manera que, si el coeficient és positiu, s'espera un augment en el nombre esperat d'èxits (en aquest cas, de crims) amb un augment en aquesta variable, i viceversa. A més dels coeficients per a les variables explicatives, el paràmetre de dispersió és crucial. Si el paràmetre de dispersió és major que 1, implica que hi ha sobredispersió en les dades i que la variabilitat és major del que s'esperaria sota una distribució de Poisson.

VI. RESULTATS

En aquest capítol, es presenten els resultats obtinguts en l'anàlisi descriptiva bivariant de les variables, juntament amb els resultats dels diferents models estimats i les interpretacions dels paràmetres corresponents.

6.1 Tractament i selecció de variables

El tractament de variables implica realitzar una fase de neteja de dades, amb l'objectiu d'identificar i corregir els possibles problemes en les dades, com valors faltants, errors en la imputació de dades o incongruències entre variables.

Primerament, es decideix excloure les dues ciutats autònomes de Ceuta i Melilla en el model per les seves característiques pròpies com també per la manca de dades disponibles per la variable **Turisme** en aquestes zones. D'aquesta manera, evitem introduir biaixos en el model que afecten al seu ajust i a la interpretació dels resultats.

Finalment, per la falta d'homogeneïtat en les dades facilitades pel Ministeri d'Interior entre els anys 2015 i 2016, comentada anteriorment en l'apartat **4.2**, es decideix no tenir-los en compte. Per tant, el període estudiat esdevé del 2017-2022.

En la construcció del model, cal escollir quines variables seran útils per a explicar la variable dependent, per consegüent, s'han seguit les següents etapes:

1. Ajustar el model saturat que conté totes les variables de la base de dades i identificar quines no són significatives.
2. Eliminar del model la variable menys significativa de totes, és a dir, que presenti un p-valor major. És important considerar alhora que, les variables es poden eliminar si compleixen condicions addicionals a no ser estadísticament significatives, com ara que el paràmetre estimat sigui similar al del model simple i que els coeficients de les altres variables no canviïn gaire a l'eliminar la variable del model.
3. A partir de la funció *anova* del paquet *stats* (R Core Team, 2020) [30], es comparen dos models, un amb totes les variables pertinents (model1) i l'altre model eliminant una d'elles, la menys significativa (model2).

Figura 4. Funció *anova* del paquet *stats* en R

```
anova(model1, model2, test='Chisq')
```

Si el p-valor d'aquest test *anova* és estadísticament no significativa (menor que el nivell de significació), vol dir que el model2 és millor que el model1, ja que amb menys variables explica millor la variable dependent.

4. Addicionalment al criteri anterior, pels models probabilístics, també s'ha tingut en compte el Criteri d'Informació Bayesiana (BIC) que es tracta d'una eina objectiva per a seleccionar models que ofereixen un bon ajust de les dades sense ser excessivament complexos. L'elecció del model amb el BIC més baix busca trobar un equilibri entre l'explicació de les dades i la simplicitat del model.

$$BIC = -2 \cdot \ln(\hat{L}) + p \cdot \ln(n)$$

6.2 Anàlisi descriptiva bivariant

Aquesta anàlisi descriptiva proporciona una visió general de la distribució i les característiques de les variables considerades, permetent una comprensió preliminar de l'estructura de les variables explicatives i la formulació dels models lineals generalitzats subsegüents. Tal com s'avançava al capítol d'anàlisi descriptiva (IV), en aquest apartat s'analitzen els factors econòmics, sòcio-culturals, demogràfics i educatius.

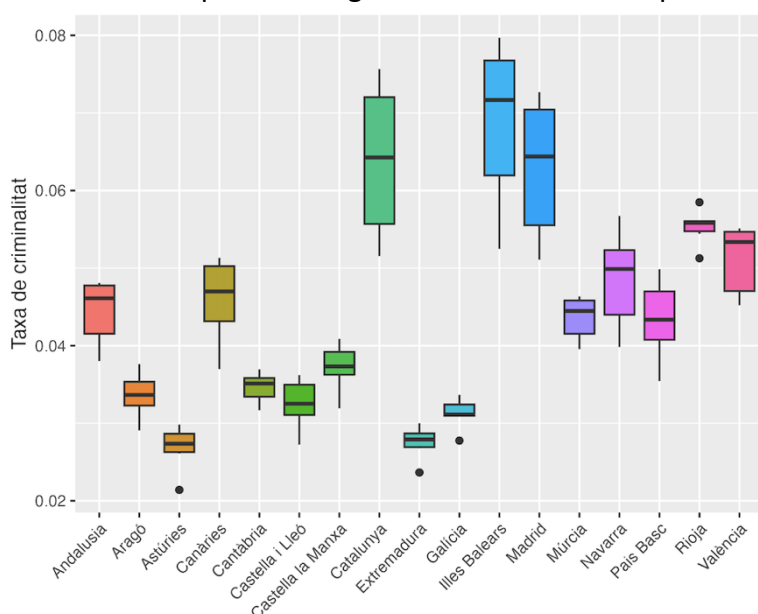
La decisió de desglossar la variable resposta de la taxa de criminalitat en dues variables és una estratègia raonada per abordar de manera més efectiva els diferents tipus de delictes i les seves raons subjacents. La primera variable incorpora els **delictes menors** com baralles, robatoris, els furtis, les sostraccions de vehicles, el tràfic de drogues i la resta de criminalitat convencional. D'altra banda, la segona variable resposta incorpora els **delictes greus** com els assassinats, intents d'assassinats, segrests i violacions. Aquest enfocament facilita la interpretació dels resultats i millora l'ajust del model, ja que permet centrar-se en categories més específiques de delictes, tenint en compte les diferències en les raons associades a cada tipus de comportament delictiu.

6.2.1 Delictes menors per CCAA

La distribució de la taxa de criminalitat⁵ per a delictes menors entre les diferents regions presenta una notable heterogeneïtat. Catalunya, illes Balears i Madrid lideren el rànquing quant a taxes més elevades, mentre que Extremadura i Astúries exhibeixen les taxes més baixes, corroborant les troballes prèvies de l'anàlisi descriptiva. A més, s'identifiquen punts extrems (*outliers*) a Astúries, Extremadura, Galícia i La Rioja, factors que podrien influir en l'ajust del model.

La posició relativa de la mediana en el *boxplot* proporciona pistes sobre la simetria o asimetria de la distribució de les dades. Per exemple, La Rioja, Cantàbria, Extremadura i Galícia presenten asimetria, indicant una concentració desigual de les dades. Catalunya i Cantàbria són les regions més simètriques i, per tant, amb una distribució més uniforme.

Gràfic 10. Comparativa Regional: Delictes menors per CCAA.



Font: Elaboració pròpia a partir de l'R i les dades del Ministeri d'Interior i l'INE.

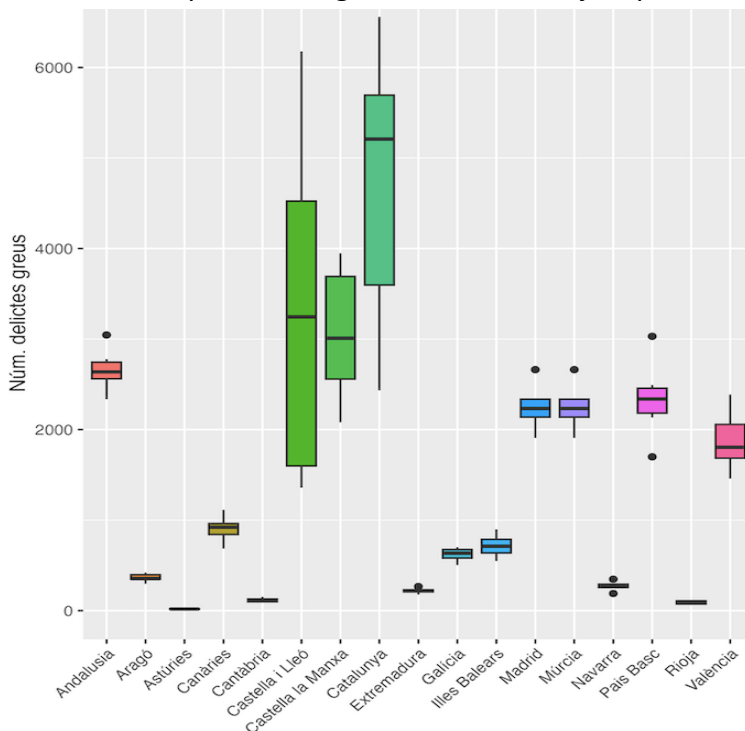
6.2.2 Delictes greus per CCAA

En analitzar els delictes greus, s'observa una imatge diferent, amb Catalunya, Castella i Lleó, i Castella-la Manxa liderant la primera posició, presentant una distribució bastant asimètrica amb bigotis bastant llargs (major dispersió), especialment en les dues primeres regions. D'altra banda, les taxes mínimes es registren a Astúries, La Rioja, Cantàbria, Extremadura i Navarra.

⁵ Incorporada en la gràfica per a facilitar la visualització, normalitza els valors significativament grans i dispers entre les regions. No obstant això, en el model s'usa el recompte directe dels delictes enlloc de la taxa.

També s'observa una major quantitat de punts extrems en comparació amb l'altre gràfic, la qual cosa suggereix una major variabilitat i dispersió en aquestes dades. Com a conseqüència d'aquesta variabilitat, podria presentar desafiaments addicionals en el procés d'ajust del model, ja que la presència de dades tan disperses pot complicar la identificació de patrons o tendències clares en la distribució.

Gràfic 11. Comparativa Regional: Delictes majors per CCAA.



Font: Elaboració pròpia a partir de l'R i les dades del Ministeri d'Interior i l'INE.

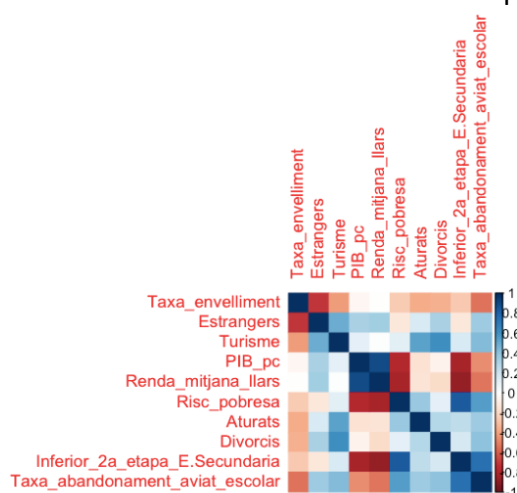
6.2.3 Matriu de correlació variables explicatives

La matriu de correlació mostra les relacions lineals entre les variables explicatives, proporcionant informació valuosa sobre les possibles relacions entre les variables abans de realitzar l'ajust del model.

Algunes de les correlacions destacades són: una correlació negativa forta entre la taxa d'envelliment i els estrangers (-0,8), indicant que a mesura que la població envella, la proporció d'estrangers tendeix a disminuir. Una altra correlació negativa forta es troba entre el PIB per càpita i el risc de pobresa, suggerint que a mesura que augmenta el PIB per càpita, la taxa de risc de pobresa tendeix a disminuir. També s'observen relacions negatives entre indicadors econòmics i la proporció de persones que tenen el nivell d'estudis inferior a la 2a etapa d'educació secundària. Totes elles representades per tons vermellosos foscos en el **gràfic 12**.

En contrast, s'observen correlacions positives fortes, representades per tons blaus fosc, entre el risc de pobresa i el fet de tenir un nivell educatiu inferior a la 2a etapa de l'educació secundària (0,7) i, d'altra banda, entre aquesta última variable i la taxa d'abandonament prematur escolar (0,7). Suggestint que una major proporció de persones amb un nivell educatiu inferior a la 2a etapa de l'educació secundària, s'associa amb una major taxa d'abandonament escolar primerenc i tendeix a augmentar el risc de pobresa.

Gràfic 12. Matriu de correlació de les variables explicatives.



Font: Elaboració pròpia a partir de l'R i les dades del Ministeri d'Interior i l'INE.

6.2.4 Variable indicadora de pandèmia

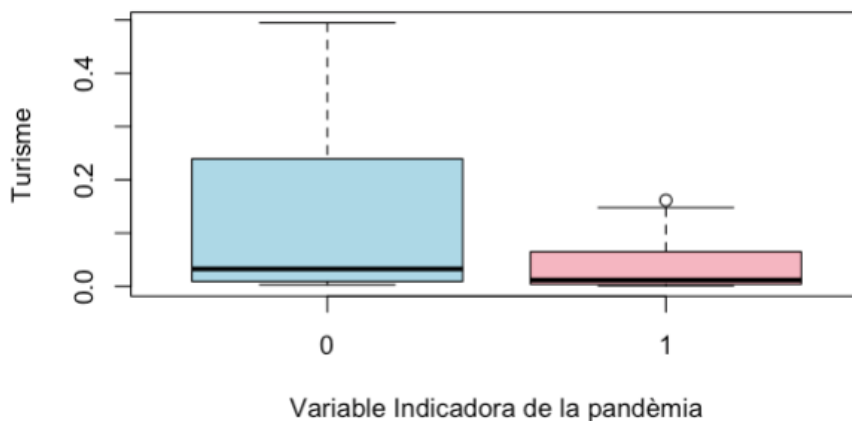
En l'anàlisi descriptiva detallada de les variables explicatives del model, s'han observat patrons inusuals al llarg dels anys (consultar l'Annex per a obtenir més detalls). Aquests valors extrems podrien atribuir-se, almenys en part, a la presència de la pandèmia de la COVID-19 durant els anys 2020-2021. En conseqüència, s'ha decidit introduir una variable indicadora binària, que rep el valor 1 si les dades corresponen als esmentats anys i, zero, en cas contrari.

Encara que la pandèmia ha tingut un efecte notable en general en totes les variables, fins i tot en les variables resposta (de reducció), s'ha decidit incloure aquella que ha experimentat un impacte més significatiu. Cal dir que, en totes les variables, s'observa una disminució dels seus valors, excepte en el cas de la presència d'estrangers, la renda mitjana per llar, el nombre d'aturats i la presència de persones amb els diferents nivells d'educació (inferior a la 2a etapa secundària, 2a etapa secundària i educació superior), que experimenten augments.

El turisme, tal com s'observa al **gràfic 13**, mostra una reducció significativa dels seus valors, explicada pel tancament i les restriccions de mobilitat aplicades a Espanya durant els anys

2020-2021. També es pot notar que, aquesta variable, presenta una distribució asimètrica, amb la mediana situada en la part inferior de la caixa, indicant que la majoria dels valors se centren en una banda baixa del rang interquartílic. A més, els bigotis són més llargs pels anys que no hi havia pandèmia, suggereint una dispersió més ampla dels valors en aquests períodes.

Gràfic 13. Boxplot de la variable indicadora de pandèmia vs. Turisme.



Font: Elaboració pròpia a partir de l'R i les dades del Ministeri d'Interior i l'INE.

6.3 Models log-lineals emprats

En aquest apartat es mostraran els diferents models estadístics ajustats per a conèixer els factors que expliquen la taxa de delictes greus i menors, juntament amb una breu interpretació dels seus paràmetres i els resultats de bondat d'ajust.

6.3.1 Model Poisson

Es pretén estudiar quins factors expliquen la taxa de criminalitat, tant pels delictes greus com pels menors. Amb aquest propòsit, s'ajusten dos models de Poisson, un amb la variable resposta *Resta_delictes* (on s'inclouen els delictes menors) i l'altre, amb la variable resposta *Delictes_contra_persona* (on s'inclouen els delictes greus). En aquests dos models s'inclouen les variables explicatives especificades en la **taula 13** del capítol 2 (metodologia).

Després de seguir les passes del llibre D.Hosmer i S.Lemeshow [11] i d'usar la funció d'R *backward stepwise* (parteix del model saturat, amb totes les variables) sota el criteri de BIC, s'obtenen els següents models log-lineals, l'expressió estimada de la qual es defineix en l'**apartat 5.1.1** del present projecte.

La inclusió del *offset* del logaritme de la població d'entre 16 i més anys és fa necessària per a obtenir la taxa de criminalitat, permetent obtenir comparacions entre les regions independentment de la seva mida poblacional.

No s'inclou en la formula de cap model la variable CCAA, ja que la variabilitat en la resposta associada amb les diferents regions està capturada per altres variables explicatives incloses en el model.

- **Model Poisson de delictes menors:**

```
glm(formula = Resta_delictes ~ offset(log(Pobl16)) + Taxa_envelliment +  
  Estrangers + Turisme + PIB_pc + Renda_mitjana_llars + Risc_pobresa +  
  Aturats + Inferior_2a_etapa_E.Secundaria + Segona_etapa_E.Secundaria +  
  E.Superior + Taxa_abandonament_aviat_escolar + Divorcis +  
  va_indicadora + Turisme * va_indicadora, family = poisson,  
  data = df)
```

- **Model Poisson de delictes greus:**

```
glm(formula = Delictes_contra_persona ~ offset(log(Pobl16)) +  
  Taxa_envelliment + Estrangers + Turisme + PIB_pc + Renda_mitjana_llars +  
  Risc_pobresa + Aturats + Inferior_2a_etapa_E.Secundaria +  
  Segona_etapa_E.Secundaria + E.Superior + Taxa_abandonament_aviat_escolar +  
  Divorcis + va_indicadora, family = poisson, data = dades)
```

Respecte a la interpretació de la **sortida 1**, totes les variables (incloent la d'interès) són significatives. La constant indica el logaritme de la taxa de delictes menors per càpita quan totes les variables explicatives del model són zero. Pel que fa a les variables explicatives, les que mostren un signe negatiu signifiquen que un increment unitari de la variable explicativa s'associa amb una disminució de la taxa de delictes menors (mantenint les altres variables constants), mentre que el signe positiu s'associa amb increments en la taxa de delictes menors. Per exemple, un increment en la taxa d'envelliment (Taxa_envelliment) s'associa amb una disminució en la taxa de delictes menors, d'altra banda, la presència de la variable indicadora de la pandèmia de la COVID-19 (va_indicadora) s'associa amb una disminució en la taxa de delictes menors. Per últim, la interacció (Turisme:va_indicadora) mostra com l'efecte del turisme canvia en presència de la variable indicadora de la pandèmia, suggerint que la relació entre turisme i delictes menors és menor quan la variable indicadora és present, és a dir, en presència de la pandèmia de la COVID-19.

En general, els signes dels coeficients semblen tenir sentit i són coherents amb la lògica esperada. Encara que no s'alineen completament amb les hipòtesis inicials que suggerien que la presència d'estrangers no tindria impacte significatiu en la delinqüència, percepció que es veu qüestionada per la significança observada en els resultats d'aquest model. És important

recordar també que la relació entre variables i delictes no implica causalitat, i unes certes troballes podrien deure's a factors no observats.

Sortida 1. Model Poisson de delictes menors

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)	
(Intercept)	1.815e+04	7.285e+02	24.91	<2e-16	***
Taxa_envelliment	-2.677e-03	2.375e-05	-112.70	<2e-16	***
Estrangers	1.946e+00	1.770e-02	109.94	<2e-16	***
Turisme	2.758e-01	4.751e-03	58.04	<2e-16	***
PIB_pc	1.727e-05	2.539e-07	68.03	<2e-16	***
Renda_mitjana_llars	-1.073e-05	2.917e-07	-36.77	<2e-16	***
Risc_pobresa	-1.136e-02	1.236e-04	-91.94	<2e-16	***
Aturats	7.318e+03	1.086e+02	67.39	<2e-16	***
Inferior_2a_etapa_E.Secundaria	-1.815e+02	7.285e+00	-24.92	<2e-16	***
Segona_etapa_E.Secundaria	-1.815e+02	7.285e+00	-24.92	<2e-16	***
E.Superior	-1.815e+02	7.285e+00	-24.92	<2e-16	***
Taxa_abandonament_aviat_escolar	1.916e-03	1.848e-04	10.37	<2e-16	***
Divorcis	1.302e-01	2.384e-03	54.60	<2e-16	***
va_indicadora	-6.132e-02	1.403e-03	-43.71	<2e-16	***
Turisme:va_indicadora	-3.645e-01	1.527e-02	-23.87	<2e-16	***

Font: Elaboració pròpia a partir de l'R i les dades del Ministeri d'Interior i l'INE.

D'altra banda, de la **sortida 2**, s'observa que totes les variables, incloent la d'interès, són significatives. La interpretació de la constant és igual a la del model anterior, indicant un nivell alt de delictes en la població basi.

En l'àmbit econòmic, un augment en el PIB per càpita (PIB_pc) s'associa a una reducció de les taxes de delictes greus. El mateix passa amb la renda mitjana per llar (Renda_mitjana_llars), indicant que regions amb majors ingressos mitjans tendeixen a tenir taxes de delictes greus més baixes.

En l'àmbit educatiu, la presència de persones amb un nivell inferior a l'educació secundària (Inferior_2a_etapa_E.Secundària) es vincula amb un augment en la taxa de delictes greus, mentre que la presència de persones amb educació superior (E.Superior) s'associa a una disminució d'aquesta. L'associació negativa de la taxa d'abandonament escolar primerenc (Taxa_abandonament_aviat_escolar) amb la taxa de delictes greus, indicant que regions amb majors taxes d'abandó escolar primerenc tendeixen a tenir taxes de delictes greus més baixes.

En l'àmbit social, la presència d'estrangers (Estrangers) i l'índex de risc de pobresa (Risc_pobresa) tenen impactes positius, indicant que estan associats a majors taxes de delictes greus. A més, el nombre de divorcis (Divorcis) mostra una relació negativa, suggerint que les regions amb més divorcis tenen taxes de delictes greus més baixes. Aquest últim resultat no acaba de concordar amb les suposicions inicials sobre els crims passionals, suggerint que altres

factors o dinàmiques podrien estar influïnt en la relació entre divorcis i delictes greus, i podria ser objecte d'estudi addicional per comprendre millor aquesta dinàmica particular.

Finalment, la variable indicadora de la pandèmia de la COVID-19 (va_indicadora) presenta un impacte negatiu en la taxa de delictes greus.

Sortida 2. Model Poisson de delictes greus

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)	
(Intercept)	6.482e+04	5.835e+03	11.109	< 2e-16	***
Taxa_envelliment	-7.936e-03	1.741e-04	-45.581	< 2e-16	***
Estrangers	3.857e+00	1.258e-01	30.663	< 2e-16	***
Turisme	1.509e-01	3.535e-02	4.269	1.96e-05	***
PIB_pc	-2.169e-05	2.115e-06	-10.257	< 2e-16	***
Renda_mitjana_llars	-5.126e-05	2.443e-06	-20.982	< 2e-16	***
Risc_pobresa	-9.312e-02	9.822e-04	-94.813	< 2e-16	***
Aturats	-4.073e+04	7.511e+02	-54.223	< 2e-16	***
Inferior_2a_etapa_E.Secundaria	-6.482e+02	5.835e+01	-11.108	< 2e-16	***
Segona_etapa_E.Secundaria	-6.483e+02	5.835e+01	-11.111	< 2e-16	***
E.Superior	-6.482e+02	5.835e+01	-11.109	< 2e-16	***
Taxa_abandonament_aviat_escolar	-1.596e-02	1.422e-03	-11.220	< 2e-16	***
Divorcis	-1.276e+00	1.843e-02	-69.214	< 2e-16	***
va_indicadora	-3.866e-01	9.244e-03	-41.821	< 2e-16	***

Font: Elaboració pròpia a partir de l'R i les dades del Ministeri d'Interior i l'INE.

Partint d'aquests dos primers models, on s'assumeix que el paràmetre de dispersió és igual a 1, cal estudiar l'existència de la sobredispersió, mitjançant el càlcul de l'estimació puntual del paràmetre de dispersió, com també el test de sobredispersió que es basa en dos contrastos d'hipòtesis per a determinar si la relació entre la variància i l'esperança és lineal (model Quasi-Poisson) o quadràtica (model Binomial Negatiu).

A través de l'estimació puntual del paràmetre de dispersió s'obtenen els següents resultats:

- **Model delictes menors:** $\hat{\phi} = 714,374$
- **Model delictes greus:** $\hat{\phi} = 482,725$

Aquest resultat ens dona indicis de que el paràmetre de dispersió no és 1, tal com assumeix el model de Poisson i que, per tant, haurem d'usar un altre model per a solucionar la problemàtica de la sobredispersió que infraestima els errors estàndards reals dels coeficients de les variables del model, fent, alhora, que molts d'ells siguin significatius.

Contrast d'hipòtesis del test de dispersió amb *trafo* = 1:

H₀: La dispersió és consistent amb la distribució de Poisson ($\theta = 1$)

$$V(Y|X) = \sigma^2 = \mu = E(Y|X)$$

H₁: Hi ha una sobredispersió real, relacionant la variància amb l'esperança de forma lineal.

$$V(Y|X) = \sigma^2 = \mu = (1 + \alpha) \cdot \mu = \phi \cdot \mu$$

Contrast d'hipòtesis del test de dispersió amb *trafo* = 2:

H₀: La dispersió és consistent amb la distribució de Poisson ($\theta = 1$)

$$V(Y|X) = \sigma^2 = \mu = E(Y|X)$$

H₁: Existeix una sobredispersió real, la variància amb l'esperança de forma quadràtica.

$$V(Y|X) = \sigma^2 = \mu + \alpha \cdot \mu^2 = \mu + \frac{1}{\theta} \cdot \mu^2$$

Pels dos models s'obtenen p-valors menors al nivell de significació (0,05), tant per *trafo* = 1 com per *trafo* = 2, el que ens porta a rebutjar la hipòtesi nul·la d'absència de sobredispersió⁶. Com a conseqüència es procedeix a ajustar un model Binomial Negatiu.

6.3.1 Model Binomial Negatiu

De la mateixa manera que amb el model de Poisson, a partir del model saturat es van traient una a una les variables que no tenen significació en el model i es va comparant el seu BIC. Obtenint els següents models:

- **Model Binomial Negatiu de delictes menors**

```
glm(formula = Resta_delictes ~ Taxa_envelliment + Estrangers +  
Turisme + Risc_pobresa + Segona_etapa_E.Secundaria + va_indicadora +  
offset(log(Pobl16)), family = negative.binomial(mod.nb01$theta),  
data = df)
```

De la **sortida 3**, s'observa que l'estimació puntual de la sobredispersió s'ha vist reduïda, en concret a 1,177, però encara és superior a 1. Per tant, encara hi ha cert grau de sobredispersió en les dades i que, com a conseqüència, la variabilitat real en les dades és major del que s'espera sobre la distribució assumida pel model. Això podria deure's

⁶ Per a un major detall de la sortida d'R, consultar els Annexos l'apartat de codis en R.

a la presència de factors no tinguts en compte en el model que contribueixen a la variabilitat addicional.

Sortida 3. Model Binomial Negatiu de delictes menors

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	-2.2885045	0.1576641	-14.515	< 2e-16 ***
Taxa_envelliment	-0.0027868	0.0005171	-5.389	5.14e-07 ***
Estrangers	2.1620681	0.3986681	5.423	4.44e-07 ***
Turisme	0.7131371	0.1027789	6.939	4.82e-10 ***
Risc_pobresa	-0.0161023	0.0017181	-9.372	3.64e-15 ***
Segona_etapa_E.Secundaria	-0.0199287	0.0066485	-2.997	0.00347 **
va_indicadora	-0.1074583	0.0249159	-4.313	3.94e-05 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for Negative Binomial(105.3976) family taken to be 1.176811)

Null deviance: 965.06 on 101 degrees of freedom
Residual deviance: 112.35 on 95 degrees of freedom
AIC: 2106.3

Number of Fisher Scoring iterations: 4

Font: Elaboració pròpia a partir de l'R i les dades del Ministeri d'Interior i l'INE.

Les possibles solucions que es donen per a corregir la sobredispersió són afegir variables explicatives que tenen influència en el model, la reconsideració d'un altre tipus de funció de distribució o l'aplicació de correccions.

A través del test de chi-quadrat de raó de versemblances entre dos models niats, es comprova si la inclusió d'alguna variable no inclosa en el model anterior té significança estadística. El resultat és negatiu, és a dir, que cap variable explicativa addicional aporta informació extra significativa al model anterior. Per això, s'opta per aplicar una altra solució. Es tracta de l'aplicació de transformacions en les variables explicatives presents. A partir dels gràfics de diagnòstic per a la validació dels models (*residualPlots(model)*) i del test *Stat*, a partir dels pvalors significants, es detecten comportaments no lineals entre la variable Turisme i el predictor lineal. Per aquest motiu, es decideix afegir la variable Turisme al quadrat, reflectint una relació quadràtica amb la variable resposta. Obtenint el següent model:

```
glm(formula = Resta_delictes ~ offset(log(Pobl16)) + Taxa_envelliment +
     Estrangers + Turisme + I(Turisme^2) + Risc_pobresa + Segona_etapa_E.Secundaria +
     va_indicadora, family = negative.binomial(mod9.nb01$theta),
     data = df)
```

A partir de la introducció de la variable Turisme al quadrat s'ha aconseguit reduir la dispersió lleugerament, reflectint la capacitat del terme quadràtic per capturar relacions no lineals associades a aquesta variable. Tot i això, és essencial tenir cura a l'hora d'introduir termes quadràtics, ja que, a banda de millorar la capacitat del model per

ajustar-se a les dades, pot afectar la interpretabilitat del model i augmentar el risc de sobre ajust. Per això, és important trobar un equilibri entre la complexitat del model i la seva capacitat de generalització a noves dades.

Sortida 4. Model Binomial Negatiu de delictes menors amb Turisme²

```

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)    -2.2116079   0.1539126  -14.369  < 2e-16 ***
Taxa_envelliment -0.0027645   0.0004979   -5.553  2.60e-07 ***
Estrangers       2.0071610   0.3879414    5.174  1.29e-06 ***
Turisme         1.5649173   0.2985028    5.243  9.71e-07 ***
I(Turisme^2)    -2.0873499   0.6880875   -3.034  0.003125 **
Risc_pobresa    -0.0175510   0.0017247  -10.176  < 2e-16 ***
Segona_etapa_E.Secundaria -0.0225917   0.0064485   -3.503  0.000705 ***
va_indicadora   -0.1038661   0.0240168   -4.325  3.80e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for Negative Binomial(105.3406) family taken to be 1.090181)

Null deviance: 964.54  on 101  degrees of freedom
Residual deviance: 102.27  on  94  degrees of freedom
AIC: 2098.3

Number of Fisher Scoring iterations: 4

Font: Elaboració pròpia a partir de l'R i les dades del Ministeri d'Interior i l'INE.

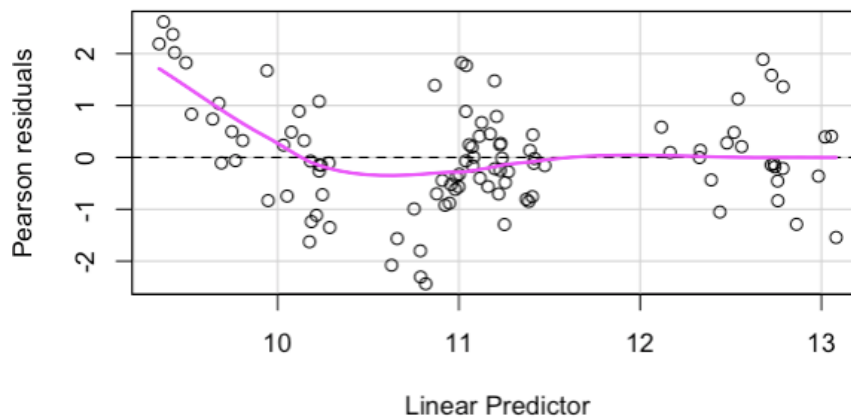
```

Es valida el model de forma numèrica com gràfica. Tant pel test de deviança com per l'estadístic de Pearson no s'observen suficients evidències estadístiques per rebutjar la hipòtesis nul·la, de manera que, el model és vàlid.

Una altra mesura que s'empra per avaluar la multicol·linealitat entre les variables predictores del model de regressió és l'anàlisi del factor d'inflació de la variància (VIF). Com totes elles presenten valors de VIF inferiors a 5, es consideren acceptables i que, per tant, no presenten multicol·linealitat greu.

A nivell gràfic la validació no seria gaire bona, sembla que el model tendeix a infraestimar els valors predits de les taxes baixes, ja que tots els residus de la part esquerra del gràfic queden per sota. La línia magenta hauria d'estar sempre molt a prop del 0 (idealment al 0). S'han emprat gairebé totes les possibles solucions per tractar per millorar la validació, però no ha estat possible trobar un millor ajust donada la limitació en l'ajust dels valors predits.

Gràfic 14. Validació gràfica del model Binomial Negatiu de delictes menors



Font: Elaboració pròpia a partir de l'R i les dades del Ministeri d'Interior i l'INE.

Independentment de la validació del model, és possible interpretar els coeficients del model. De la constant, trobem un valor negatiu, indicant que s'espera una taxa de criminalitat baixa en absència de les altres variables explicatives. L'odds ràtio (OR) associada a l'intercepta és 0,1095 ($e^{-2.211}$). Això indica que quan totes les altres variables són zero, s'espera que les probabilitats d'ocurrència d'un delictes menor siguin aproximadament 0,1095 vegades les probabilitats de no ocurrència.

Pel que fa a la taxa d'envelliment, l'OR associat a la taxa d'envelliment és de 0,9972 ($e^{0.028}$). Això suggereix que, per cada unitat d'augment en la taxa d'envelliment, les probabilitats d'ocurrència d'un delictes menor disminueixen en aproximadament un 0,28%.

Respecte als estrangers i el turisme, segons l'últim model estimat, el seu OR ens indica que hi ha una relació positiva en la taxa de criminalitat de delictes menors.

La variable quadràtica de turisme captura la relació no lineal entre turisme i la taxa de criminalitat de delictes menors. L'OR d'aquesta variable és menor a 1 (0,1240), indicant que l'efecte de la variable quadràtica està associat amb una disminució en les probabilitats d'ocurrència d'un delictes menor. Per tant, inicialment la variable turisme presenta un augment en la taxa, però a mesura que aquesta variable creix, l'efecte pot veure's disminuït.

L'odds rati de la variable risc de pobresa, ens suggereix que per cada unitat d'augment en aquesta variable, les probabilitats d'ocurrència d'un delictes menor disminueixen en aproximadament un 1,74%. Un comportament semblant experimenta el número de persones amb la segona etapa d'educació secundària, que per cada unitat d'augment

en aquesta variable, les probabilitats d'ocurrència d'un delict menor disminueixen en aproximadament un 2,23%.

Finalment, el coeficient de la variable indicadora ens suggereix que les regions en els anys 2020-2021 podrien tenir una taxa de criminalitat més baixa.

- **Model Binomial Negatiu de delictes greus**

```
glm(formula = Delictes_contra_persona ~ Taxa_envelliment + Risc_pobresa +
     Segona_etapa_E.Secundaria + E.Superior + Divorcis + offset(log(Pobl16))),
     family = negative.binomial(mod2.nb01$theta), data = dades)
```

Es segueix el mateix procediment que en el model Binomial Negatiu dels delictes menors. A partir de la **sortida 5**, es pot observar que l'estimació puntual de la sobredispersió es redueix considerablement (1,378), però encara superior de l'1. S'ha d'intentar apropar al màxim aquesta estimació a l'1, ja que la presència de sobredispersió infraestima els errors estàndards de les variables del model.

Sortida 5. Model Binomial Negatiu de delictes greus

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)	
(Intercept)	4.453833	1.850430	2.407	0.018000	*
Taxa_envelliment	-0.013513	0.002427	-5.569	2.33e-07	***
Risc_pobresa	-0.082530	0.019922	-4.143	7.39e-05	***
Segona_etapa_E.Secundaria	-0.132187	0.047461	-2.785	0.006446	**
E.Superior	-0.067815	0.018503	-3.665	0.000405	***
Divorcis	-1.448060	0.363020	-3.989	0.000130	***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for Negative Binomial(2.0303) family taken to be 1.377799)

Null deviance: 170.07 on 101 degrees of freedom
Residual deviance: 121.29 on 96 degrees of freedom
AIC: 1597.9

Number of Fisher Scoring iterations: 14

Font: Elaboració pròpia a partir de l'R i les dades del Ministeri d'Interior i l'INE.

A partir dels gràfics de diagnòstic per a la validació dels models (*residualPlots(model)*) i el test *Stat*, a partir dels pvalors significants, es detecten comportaments no lineals entre la variable *Segona_etapa_E.Secundaria* i el predictor lineal, per una banda, i la variable *E. Superior* i el predictor lineal, per l'altra banda. Per aquest motiu, es decideix afegir la variable *Segona_etapa_E.Secundaria* i *E. Superior* al quadrat, reflectint una relació quadràtica amb la variable resposta. Quan es torna a comprovar si hi ha alguna altra relació no lineal, s'observa que la variable *E. Superior*² és molt significativa en el test *Stat*, indicant-nos que cal afegir un terme cúbic a aquesta variable. A partir d'aquí, es

realitza la prova de versemblança (*Likelihood Ratio test*) dels dos models, com també es calcula els seus BIC corresponents. El resultat ens porta a obtenir el següent model:

```
glm(formula = Delictes_contra_persona ~ offset(log(Pobl16)) +
     Taxa_envelliment + Risc_pobresa + Segona_etapa_E.Secundaria +
     I(Segona_etapa_E.Secundaria^2) + E.Superior + I(E.Superior^2) +
     I(E.Superior^3) + Divorcis, family = negative.binomial(mod2.nb11$theta),
     data = dades)
```

A partir de la introducció del terme quadràtic en la variable Segona_etapa_E.Secundaria i el terme cúbic en la variable E. Superior, s'aconsegueix reduir la dispersió lleugerament (1,164), com també ajuda a explicar millor la variabilitat en la variable resposta.

Encara que la variable Segona_etapa_E.Secundària juntament amb el seu terme quadràtic deixen de ser significatius individualment amb la introducció del terme cúbic de la variable E.Superior, es decideix mantenir-les perquè la seva inclusió millora l'ajust global del model, segons el BIC i el test chi-quadrat de l'anova.

Sortida 6. Model Binomial Negatiu de delictes greus amb termes quadràtics i cúbics

```
Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)    -4.755e+01  9.744e+00  -4.880 4.37e-06 ***
Taxa_envelliment -1.341e-02  2.135e-03  -6.278 1.08e-08 ***
Risc_pobresa    -9.195e-02  1.703e-02  -5.399 5.12e-07 ***
Segona_etapa_E.Secundaria  1.155e+00  7.145e-01  1.616 0.109409
I(Segona_etapa_E.Secundaria^2) -2.832e-02  1.557e-02  -1.819 0.072149 .
E. Superior      2.793e+00  6.646e-01  4.202 6.07e-05 ***
I(E.Superior^2)   -7.094e-02  1.578e-02  -4.494 2.01e-05 ***
I(E.Superior^3)   5.688e-04  1.230e-04  4.623 1.22e-05 ***
Divorcis         -1.254e+00  3.131e-01  -4.004 0.000125 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for Negative Binomial(2.4574) family taken to be 1.16436)

Null deviance: 205.43 on 101 degrees of freedom
Residual deviance: 109.07 on 93 degrees of freedom
AIC: 1571

Number of Fisher Scoring iterations: 12
```

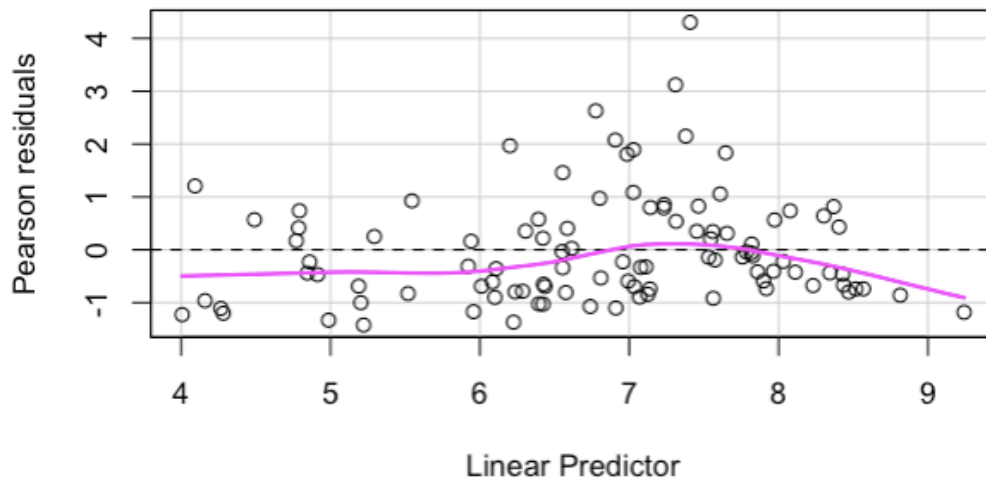
Font: Elaboració pròpia a partir de l'R i les dades del Ministeri d'Interior i l'INE.

De la mateixa manera que anteriorment, es procedeix a la validació del model, tant numèrica com gràfica. L'escena és diferent, les dues mesures numèriques es contradiuen, el p-valor de l'estadístic de Pearson surt significatiu, mentre que el test de la deviança no ho és. Aquest fet es deu a que l'estadístic de Pearson pot no comportar-se bé amb mostres petites, mentre que el test de deviança és robust en mostres petites.

A nivell gràfic, s'observa una curvatura que podria indicar una possible violació de la suposició d'homoscedasticitat, on s'espera que la variabilitat dels errors sigui constant al llarg del predictor lineal. Per a abordar aquesta problemàtica es consideren aplicar

transformacions en les variables predictors per a la obtenció d'una relació més lineal, però no es troba cap que la millori. També s'examinen la presència de punts extrems (*outliers*) i influents, on s'observen varies observacions influents amb residus grans que podrien estar afectant a l'ajust del model. Es decideix mantenir totes les observacions ja que la seva eliminació tampoc millora la validació.

Gràfic 15. Validació gràfica del model Binomial Negatiu de delictes greus



Font: Elaboració pròpia a partir de l'R i les dades del Ministeri d'Interior i l'INE.

Independentment de la validació del model, és possible interpretar els coeficients del model a partir de l'aplicació de la funció exponencial a cada coeficient del model (OR). De la constant, podem afirmar que les probabilitats de delictes greus són significativament més baixes quan les variables explicatives són 0.

Respecte les variables explicatives, la taxa d'envelliment mostra que un augment de l'1% en la seva taxa, està associat amb una disminució del 1,33% en les probabilitats de delictes greus. Si ens fixem en la variable econòmica del risc de pobresa, indica que un increment d'una unitat, en aquesta variable, es relaciona amb una disminució del 8,78% en les probabilitats de delictes greus.

Respecte a les variables educatives, s'observa que el coeficient de la variable Segona_etapa_E.Secundaria no és significatiu a un nivell de confiança del 95%, indicant que no té un efecte significatiu en la taxa de delictes greus. El seu terme quadràtic tampoc assoleix la significació del 95%, mostra una tendència significativa a un nivell del

90%, suggerint una possible relació quadràtica amb la taxa de delictes greus. D'altra banda, la quantitat de persones amb educació superior presenta un impacte substancial, ja que un augment d'una unitat està associat a un increment significatiu en les probabilitats de delictes greus. A més, la inclusió de termes quadràtics i cúbics en aquesta última variable proporciona detalls addicionals sobre la complexitat de la seva relació amb la incidència de la criminalitat. El terme quadràtic suggereix una possible atenuació en l'impacte a mesura que la proporció d'educació superior augmenta, i el terme cúbic indica un augment mínim després d'un cert punt, potser indicant un punt de saturació en aquesta relació. Cal assenyalar que, aquestes interpretacions estan subjectes a l'entorn específic del model utilitzat i els seus límits en termes de causalitat.

VII. DISCUSSIÓ I CONCLUSIONS

D'entrada, és important assenyalar que, malgrat que la teoria subjacent a l'obtenció i neteja de les dades pot semblar relativament senzilla, la realitat presenta una complexitat major. Si és veritat que les dades són extretes de les estadístiques del Ministeri d'Interior i l'INE, que són organismes estretament vinculats en la seva activitat i que el contrast d'informació provinent d'aquestes dues fonts es considera fiable. No obstant això, és important reconèixer que, en moltes ocasions, l'homogeneïtat en les dades pot ser un desafiament, la qual cosa complica el seu tractament i posterior anàlisi. De la mateixa manera que acceptem la existència de biaix en les dades i la presència de dades fosques en les estadístiques oficials sobre la delinqüència.

Abans de desenvolupar els model estadístics, es va dur a terme un anàlisi descriptiva de les taxes de criminalitats anuals i trimestrals, així com del nombre de tipologies penals incloses en aquestes taxes. Aquesta anàlisi tenia com a objectiu destacar la diversitat de patrons delictius com també la seva distribució geogràfica, suggerint que factors com el turisme, variables socioeconòmiques i l'estacionalitat influeixen en els nivells de criminalitat a les diferents regions espanyoles. De fet, es van identificar característiques comunes en les regions amb majors (Catalunya, Madrid i les Illes Balears) i menors taxes de criminalitat (Astúries, Extremadura i La Rioja). A més, s'ha observat que les dues ciutats autònomes, Ceuta i Melilla, requereixen d'un anàlisi diferenciat de la resta de les regions per les seves característiques polítiques, socials i geogràfiques pròpies que porten a tenir un comportament delictiu distintiu, amb freqüència liderant en diverses tipologies penals (robatoris, tràfic de drogues).

Respecte els models estadístics, s'opta per usar el model binomial negatiu, ja que és el més adequat per a explicar els factors que influeixen en la taxa de criminalitat. Ara bé, per l'obtenció d'un anàlisi més precís, es segmenta la variable resposta en dos grups de delictes. El primer grup, denominat com a delictes menors, engloba una àmplia gamma de crims com robatoris, baralles, furts, tràfic de drogues i altres delictes convencionals. El segon grup, classificat com a delictes greus, inclou aquells que afecten directament a la persona d'una manera més greu, com ara assassinats, intents d'assassinats, segrests i agressions sexuals.

Model de delictes menors

D'aquest model, s'observen diverses variables estadísticament significatives que influeixen en aquest tipus de delicte. Les troballes han estat:

- 1- Les **regions** amb una **població més envellida**, experimenten **menys delictes menors**, indicant una possible relació inversa entre l'edat de la població i els delictes menors.
- 2- Les **àrees** amb una **major població d'estrangers** podrien experimentar una **major incidència de delictes menors**. La significació de la variable estrangers suggereix la importància d'una comprensió més profunda del perfil socioeconòmic i educatiu d'aquest grup ja que per la falta d'aquesta informació es limita la capacitat de discernir els factors subjacents que contribueixen a la seva significació.
- 3- **Inicialment el turisme incrementa la taxa de criminalitat de delictes menors**, però a mesura que el nombre de **turismes creix**, **l'efecte es redueix**. Mostrant una dinàmica complexa entre el turisme i la taxa de criminalitat de delictes menors en les regions estudiades.
- 4- Les **regions amb més risc de pobresa** experimenten **menys delictes menors**. Sembla contradir una de les hipòtesis inicials que plantejava que una situació econòmica desafavorida podria incentivar la comissió de delictes. Això podria explicar-se pel fet que, en regions amb majors índexs de pobresa, les persones podrien tenir menys recursos per a l'activitat delictiva o altres factors socioeconòmics podrien influir en la prevalença d'aquests delictes d'una manera que no es pot capturar directament pel model.
- 5- Les **àrees** amb un nombre **més** elevat de **població amb educació secundària**, presenten una **taxa de criminalitat de delictes menors** lleugerament **inferior**. Aquesta troballa concorda amb la hipòtesi inicial que suggeria que una menor educació podria relacionar-se amb una major incidència en l'execució de delictes.
- 6- La **pandèmia de la COVID-19** ha tingut un impacte significatiu en la **reducció** de la taxa de criminalitat pels **delictes menors**.

Model de delictes greus

D'aquest model, s'observen diverses variables estadísticament significatives que influeixen en aquest tipus de delicte. Les troballes han estat:

- 1- Les **regions** amb una **població més envellida** mostren una tendència a tenir una **menor incidència de delictes greus**, suggerint una possible relació inversa entre l'edat de la població i aquest tipus de delictes.
- 2- Les **zones amb un major risc de pobresa** presenten una **menor incidència de delictes greus**. Aquest resultat contradiu la hipòtesi inicial, prèvia a l'anàlisi de les dades, que plantejava que una situació econòmica desafavorida podria incentivar la comissió d'aquests delictes.
- 3- En general, les regions amb un **major nombre de persones amb educació superior** presenten **més probabilitats de registrar delictes greus**. No obstant això, s'observa una **relació no lineal**, indicant que, després d'un **cert nivell d'educació superior**, **l'impacte** en l'augment de les probabilitats de delictes greus és molt lleuger o fins i tot es **redueix lleugerament**. Per tant, un increment moderat en el nombre de persones amb educació superior, podria mitigar les probabilitats de delictes greus fins a cert límit.
- 4- Les **regions** amb un **major nombre de divorcis** tendeixen a tenir **menors probabilitats** d'experimentar **delictes greus**. Aquesta relació podria suggerir la influència de factors sociofamiliars en els nivells de criminalitat en aquest àmbit.

A partir d'aquests resultats, és possible concloure que els factors que expliquen la taxa de criminalitat varien segons el tipus de delicte analitzat. Pels delictes menors, els factors significants són l'edat de la població, la presència d'estrangers, l'impacte del turisme, el nivell de risc de pobresa i el nivell educatiu de la població. En contrast, pels delictes greus, els factors destacats són l'edat de la població, el risc de pobresa, el nivell educatiu superior i el nombre de divorcis.

És important ressaltar que, aquests resultats, es deriven d'una anàlisi estadística i que la causalitat no pot establir-se definitivament. La complexitat de la relació entre les variables i la criminalitat suggereix la possible influència d'altres factors no considerats en el model. En aquest context, seria enriquidor explorar variables psicològiques dels individus, ja que podrien

proporcionar una comprensió més profunda dels factors subjacents a la criminalitat. No obstant això, la falta d'accés a aquestes dades en fonts públiques, motivada per qüestions de privacitat i protecció de dades, ha limitat aquesta consideració en l'anàlisi actual.

Finalment, seguint les recomanacions del professor J. Cortès en Models Lineals Generalitzats, es planteja com a projecte futur l'aplicació d'un model lineal mixt generalitzat amb efectes aleatoris (model no estudiat durant el grau).

BIBLIOGRAFIA

- [1] Anwar, A., Arshed, N., & Anwar, S. (2017). Socio-economic determinants of crime: An empirical study of Pakistan. *International Journal of Economics and Financial Issues*, 7(1), 312-322. <https://econjournals.com/index.php/ijefi/article/view/3504>
- [2] Asociados, I. (2022, 11 octubre). *El mayor problema d'Espanya son els immigrants, ¿o no?* Isacura Asociados. <https://isacuraasociados.com/blog/el-mayor-problema-de-espana-son-los-inmigrantes-o-no/>
- [3] Barreto, M. (2002). "Deserción escolar y criminalidad". División de Seguridad Ciudadana, Ministerio del Interior, Gobierno de Chile. Serie Análisis. Nº1.
- [4] Cáceres Bautista, N. A., Farro Salinas, N. T., Uribe Aliaga, C. E., & Alberto Chamorro, K. P. (2022). Què determina la criminalitat en els països? Un anàlisi economètric. *Desafiaments: Economia Y Empresa*, 2(002), 75-84. <https://doi.org/10.26439/ddee2022.n002.5712>
- [5] Cardona Isaza, (2020). Factors implicats en el fracàs escolar d'adolescents amb conducta delictiva. *Revista Criminalitat.*, vol. 62, no. 2. ISSN 1794-3108.
- [6] Carretero, N., Carretero, N., & Carretero, N. (2022, 16 octubre). La guerra de Ceuta: "Estan convertint a xavals en sicaris". *El País*. <https://elpais.com/espana/2022-10-16/la-guerra-de-ceuta-estan-convirtiendo-a-chavales-en-sicarios.html>
- [7] Carvajal Quintana, A. J., & Rincón Martínez, A. K. (2016). Influència de las pautas de criança en joves infractores com factors determinants en la comissió de conductes antisocials i delictives.
- [8] César Pol, J., & Silvestrini, R. (2004). Crim i economia subterrània a Puerto Rico. *Revista de ciències socials*, 13, 106–131. <https://revistas.upr.edu/index.php/rcs/article/view/5896>

- [9] Cid Moliné, J. (2017). *Anàlisi ecològic de la delinqüència en la ciutat de Barcelona*. Universitat Autònoma de Barcelona (UAB).
- [10] Corman, H. & Mocan, N. (2005). "A Time-Series Analysis of Crime, Deterrence, and Drug Abuse in New York City". En: *The American Economic Review*, 90: 584 – 604.
- [11] Hosmer, D., Lemeshow, S. (2000). *Applied Logistic Regression* (2nd ed.). John Wiley & Sons.
- [12] Estévez-Calapaqui, N. (2021). Relació entre desocupació i delinqüència. *Revista Científica Saberes 5.0*, 1(2), 38–47. Recuperat de <https://revistas.saberescincopuntocero.com/index.php/rcs50/article/view/111>
- [13] García España, E. (2019). Més immigració, menys delicte. *Crítica penal i poder*, (Exemplar dedicat a: La mobilitat humana: entre els drets i la criminalització), 194-205. ISSN-e 2014-3753.
- [14] García Ramos, J. A., Ramos González, C., y Ruiz Garzón, G. (2009). *Estadística empresarial*. Cádiz: Servicio de Publicaciones de la Universidad de Cádiz.
- [15] González, À. (2021). *Tasa de criminalidad en España (Estudio de la Convergencia)* - *Universidad de Zaragoza Repository*. <https://zaguan.unizar.es/record/109787#>
- [16] Fernández Bessa, C., (2010). Mobilitat sota sospita. El convenient vincle entre immigració i criminalitat en les polítiques migratòries de la Unió Europea. REMHU - *Revista Interdisciplinaria de Mobilitat Humana*, 18(35), 137-154.
- [17] Gerstenblith, Patty, *La Destrucció del Patrimoni Cultural: Un Crim Contra la Propietat o un Crim Contra les Persones?* (31 de maig de 2016). *John Marshall Review of Intellectual Property Law*, Vol. 15, No. 336, 2016, Disponible a SSRN: <https://ssrn.com/abstract=2787347>

- [18] Heller, S. B., Jacob, B. A., & Ludwig, J. (2011). *Control del delito: Estrategias i compromisos: Ingresos familiares, Pobreza veinal i Crim* (Vols. 419 –459).
<https://www.sciencedirect-com.sire.ub.edu/science/article/pii/S0049089X12000026?via%3Dihub>
- [19] INE - Instituto Nacional de Estadística. (s. f.). *Productos y servicios / Publicaciones / Publicaciones de descarga gratuita*.
https://www.ine.es/ss/Satellite?c=INESeccion_C&p=1254735110672&pagename=ProductosYServicios%2FPYSLayout&cid=1259945060213&L=0#:~:text=A%20nivel%20nacional%2C%20la%20tasa,un%20valor%20de%2048%2C6
- [20] Ivashkiv, A. (2020a, octubre 6). *Análisis de la criminalidad en las comunidades autónomas de España*. Enter Data. <https://entredatos.es/analisis-de-la-criminalidad-en-las-comunidades-autonomas-de-espana/>
- [21] Ivashkiv, A. (2020b, octubre 6). *Análisis de la criminalidad en las comunidades autónomas de España*. Enter Data. <https://entredatos.es/analisis-de-la-criminalidad-en-las-comunidades-autonomas-de-espana/#post-1269-41mghml>
- [22] La Criminalidad. (2013). *Revista Jurídica*, 85-98.
<https://core.ac.uk/download/pdf/236361479.pdf>
- [23] Lochner, L. (2008). Education and crime. *International Encyclopedia of Education*.
- [24] Machin, S., Marie, O., & Vujić, S. (2011). THE CRIME REDUCING EFFECT OF EDUCATION. *The Economic Journal*, 121(552), 463–484.
<http://www.jstor.org/stable/41236987>
- [25] Montolio, Daniel & Planells, Simón. (2012). Estimula el turisme l'activitat criminal? Evidència per a les províncies espanyoles.

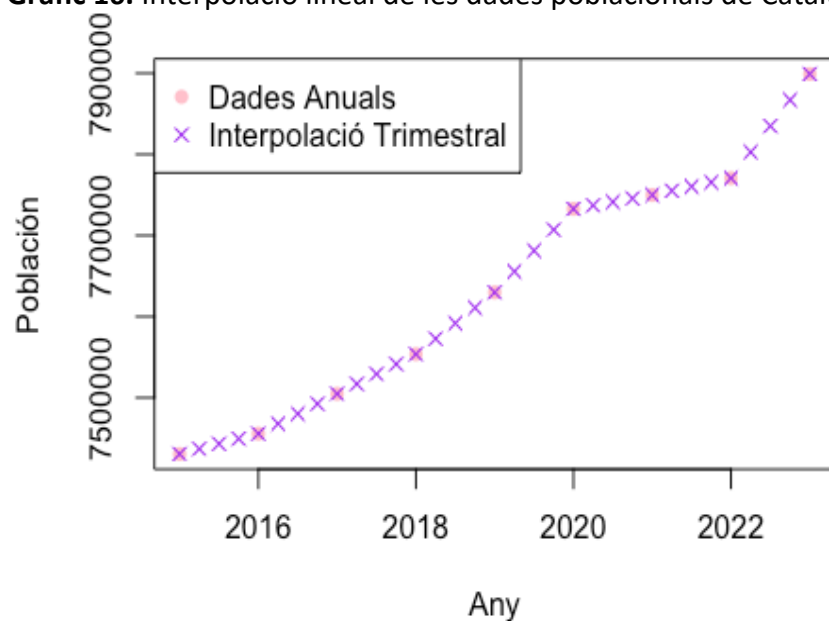
- [26] Morales, M. L. R. (2018). *Factores determinantes de la criminalidad en España. Estudio de la correlación estadística*.
<https://www.redalyc.org/journal/2739/273960279003/html/>
- [27] Navarrete Escobedo, D. (2014). Tourism and crime. Critical approach from a case study. Turismo y delito. Aproximación crítica a partir de un estudio de caso. Palapa (Revista de investigacion cientifica en arquitectura), II (2), 92-104. ISSN: 1870-7483.
Recuperado de:
https://www.academia.edu/en/36313101/Tourism_and_crime_Critical_approach_from_a_case_study_Turismo_y_delito_Aproximaci%C3%B3n_cr%C3%ADtica_a_partir_de_un_estudio_de_caso
- [28] Phillips, J., & Land, K. C. (2012). The link between unemployment and crime rate fluctuations: An analysis at the county, state, and national levels. Social Science Research, 41(3), 681-694. <https://doi.org/10.1016/j.ssresearch.2012.01.001>
- [29] Ramírez de Garay, Luis David. (2014). Crim i economia: una revisió crítica de les explicacions econòmiques del crim. Arguments (Mèxic, D.F.), 27(74), 261-290.
Recuperat el 14 de novembre de 2023, de
[http://www.scielo.org.mx/scielo.php?script=sci_arttext&pid=S0187-57952014000100010&lng=es&tlng=es.](http://www.scielo.org.mx/scielo.php?script=sci_arttext&pid=S0187-57952014000100010&lng=es&tlng=es)
- [30] R Core Team. (2021). R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from
<https://www.R-project.org/>.
- [31] Ruiz, M. (2018). Factors determinants de la criminalitat a Espanya. Estudi de la correlació estadística. Vía Iuris, (25), 55-76.

- [32] Sánchez Arteaga, T. del S. (2021). El enfoc psicològic en la criminalística: Criminologia i ciències forenses.
- [33] Téllez, J. T., & Soler, A. M. (2022). Delinqüència juvenil a Espanya: Un anàlisi empíric després de la crisi econòmica de 2008. *Revista de ciències socials*, 28(1), 20-32.
- [34] Uceda-Maza, Francesc X., & Domínguez Alonso, Javier. (2017). Vinculación entre la vulnerabilidad y la exclusión social y las trayectorias delictivas. Un estudio de asociación. *Psychosocial Intervention*, 26(1), 29-37.
<https://dx.doi.org/10.1016/j.psi.2016.07.003>
- [35] Varona Martínez, G. (2012). Anàlisi espacial de dades georreferencials d'interès criminològic en la CCAA d'Euskadi. *Kriminologiaren Euskal Institutua*.
<https://www.ehu.eus/documents/1736829/2153006/Analisis+espacial+de+datos+georreferenciales+de+interes+criminologico.pdf>
- [36] *Vista de más inmigración, menos delincuencia*. (s. f.).
<https://revistes.ub.edu/index.php/CriticaPenalPoder/article/view/30493/30700>
- [37] Vozmediano, L. y San Juan, C. (2010): *Criminologia Ambiental: ecologia del delictes i de la seguretat*. Barcelona: Editorial UOC.

ANNEX

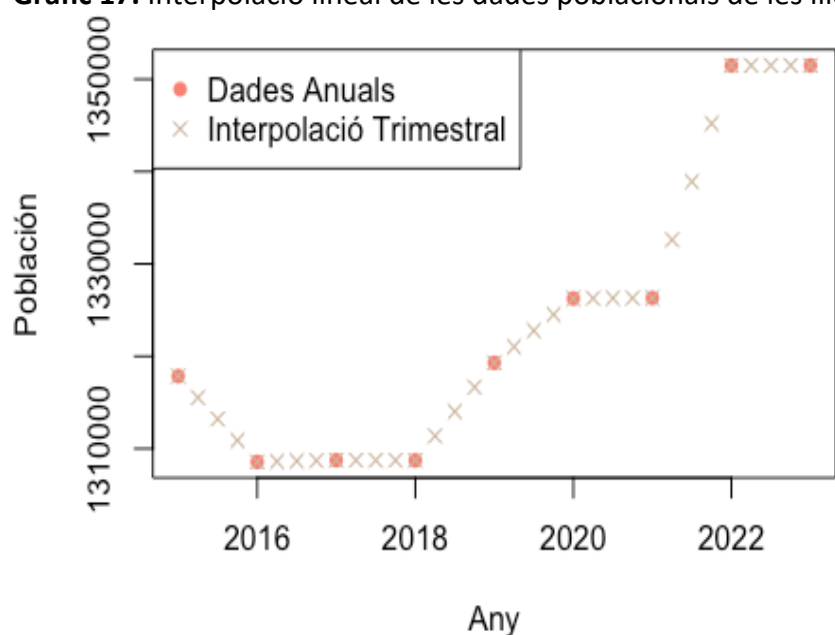
A. Gràfics complementaris

Gràfic 16. Interpolació lineal de les dades poblacionals de Catalunya



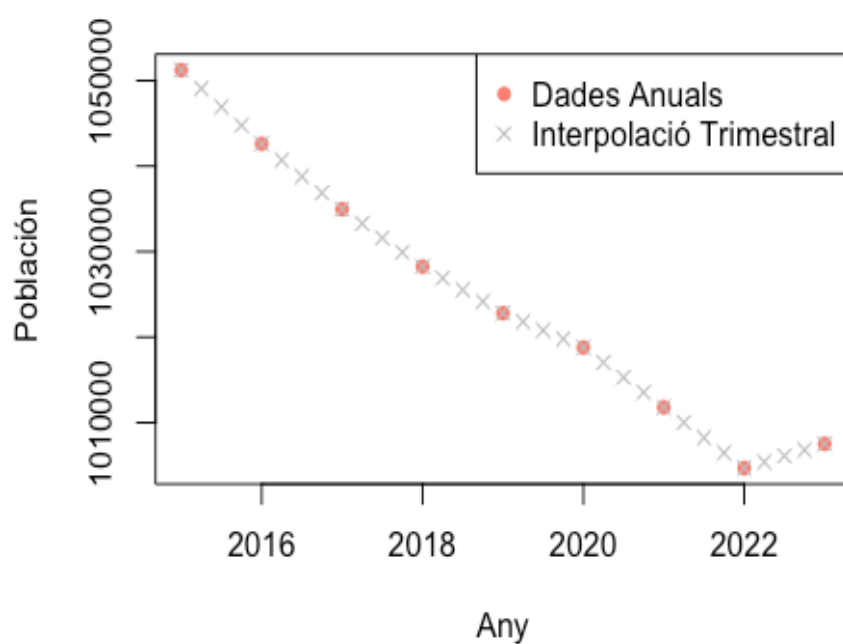
Font: Elaboració pròpia a partir de l'R i les dades de l'INE.

Gràfic 17. Interpolació lineal de les dades poblacionals de les Illes Balears



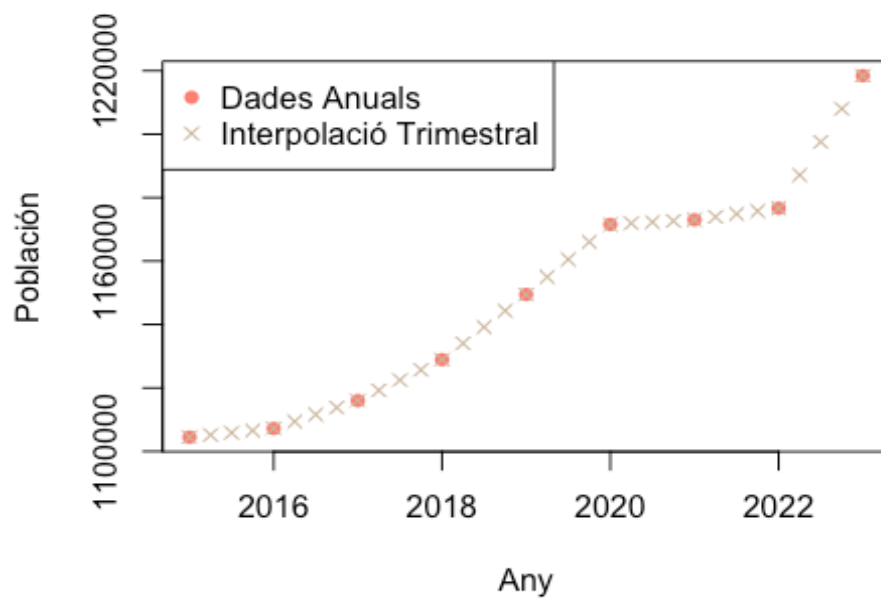
Font: Elaboració pròpia a partir de l'R i les dades de l'INE.

Gràfic 18. Interpolació lineal de les dades poblacionals d'Astúries



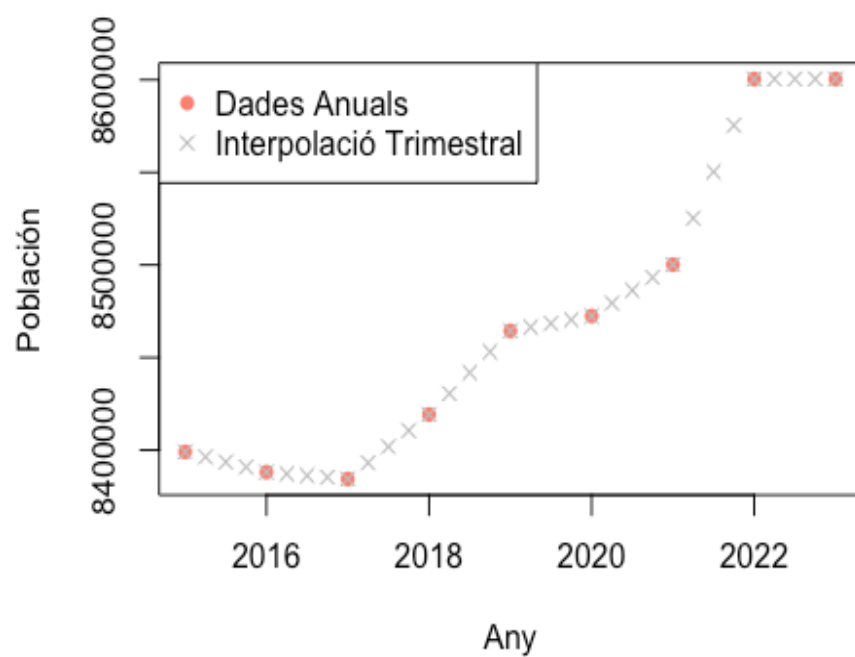
Font: Elaboració pròpia a partir de l'R i les dades de l'INE.

Gràfic 19. Interpolació lineal de les dades poblacionals d'Aragó



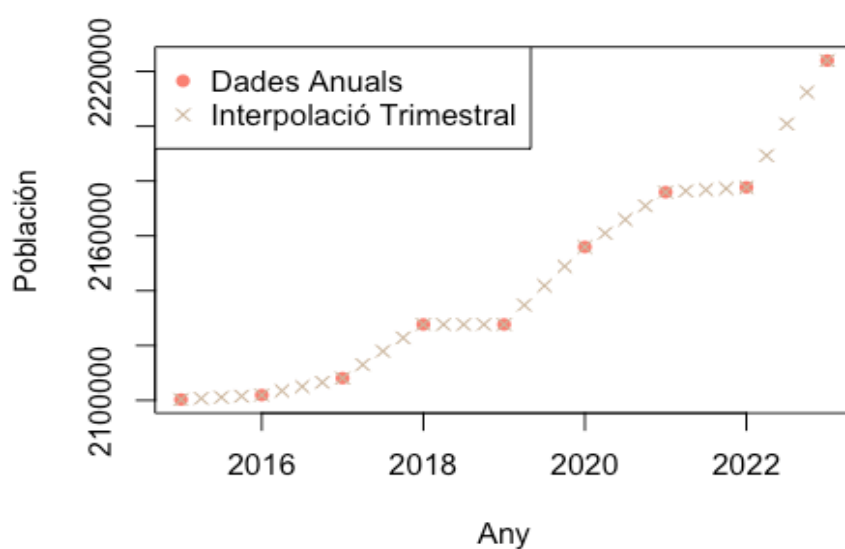
Font: Elaboració pròpia a partir de l'R i les dades de l'INE.

Gràfic 20. Interpolació lineal de les dades poblacionals d'Andalusia



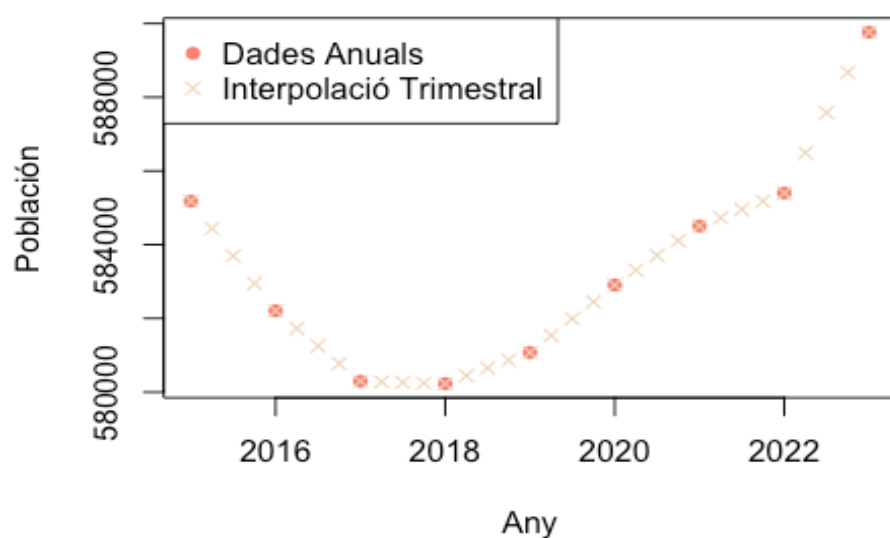
Font: Elaboració pròpia a partir de l'R i les dades de l'INE.

Gràfic 21. Interpolació lineal de les dades poblacionals de Canàries



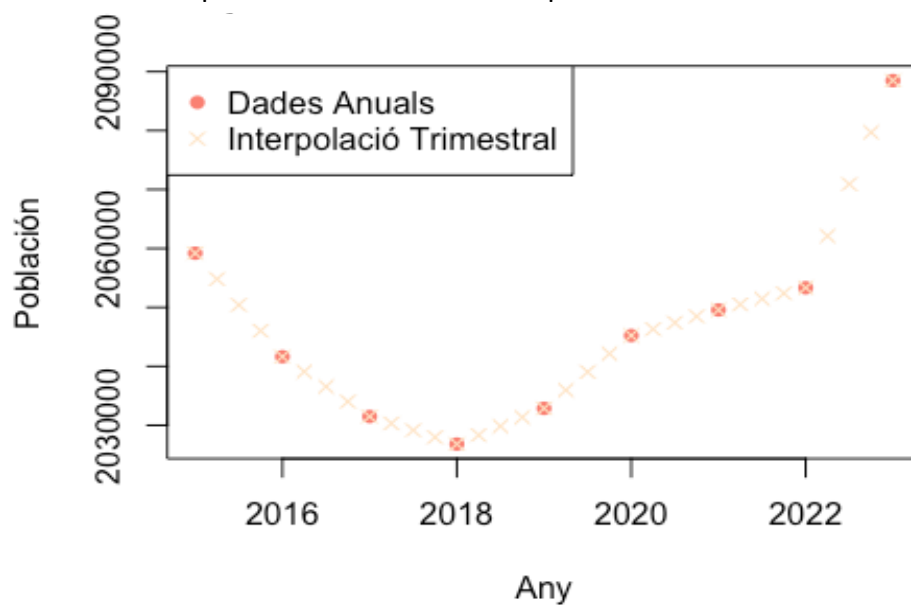
Font: Elaboració pròpia a partir de l'R i les dades de l'INE.

Gràfic 22. Interpolació lineal de les dades poblacionals de Cantàbria



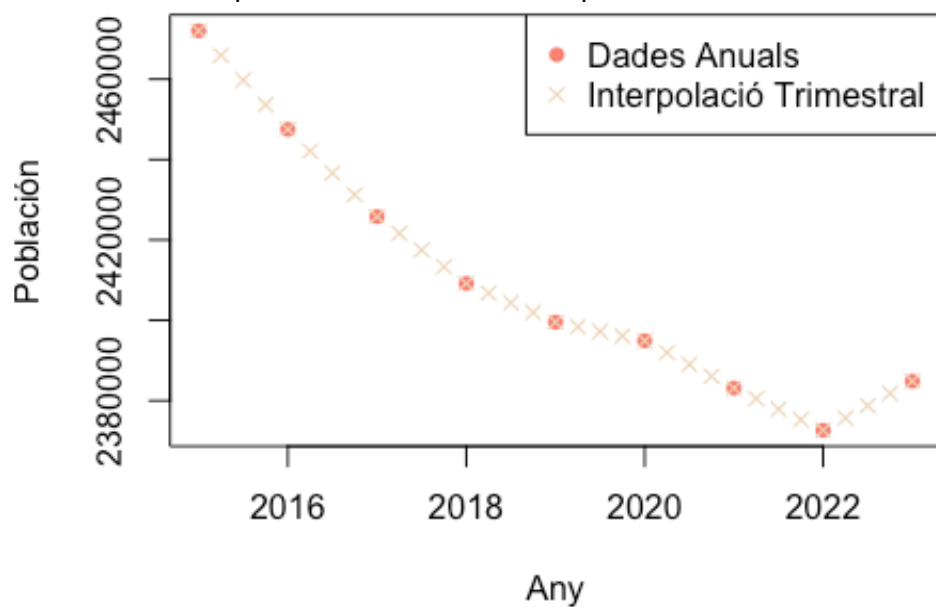
Font: Elaboració pròpia a partir de l'R i les dades de l'INE.

Gràfic 23. Interpolació lineal de les dades poblacionals de Castella la Manxa



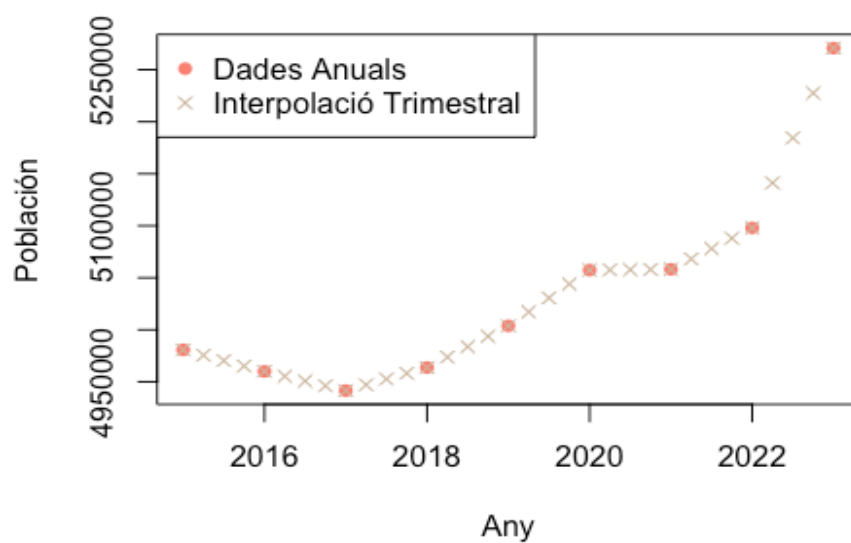
Font: Elaboració pròpia a partir de l'R i les dades de l'INE.

Gràfic 24. Interpolació lineal de les dades poblacionals de Castella i Lleó



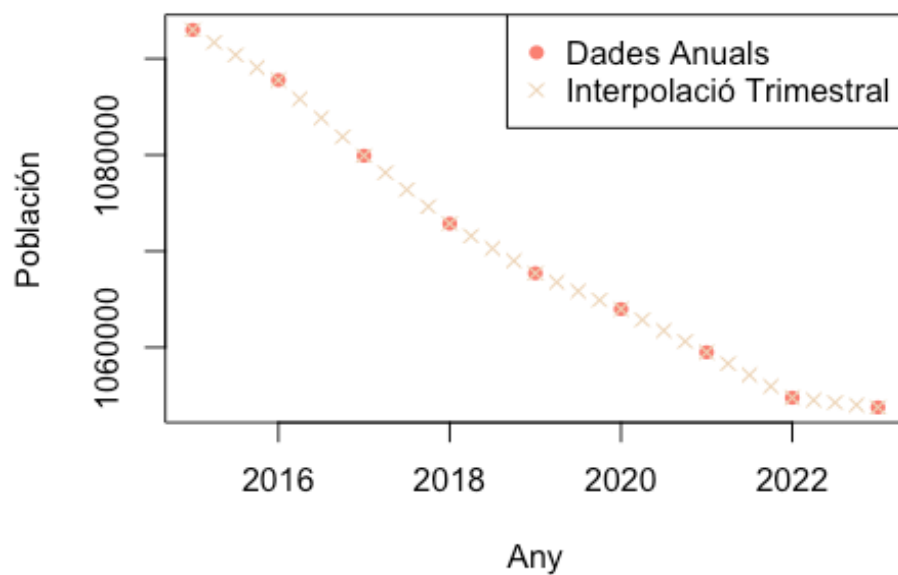
Font: Elaboració pròpia a partir de l'R i les dades de l'INE.

Gràfic 25. Interpolació lineal de les dades poblacionals de València



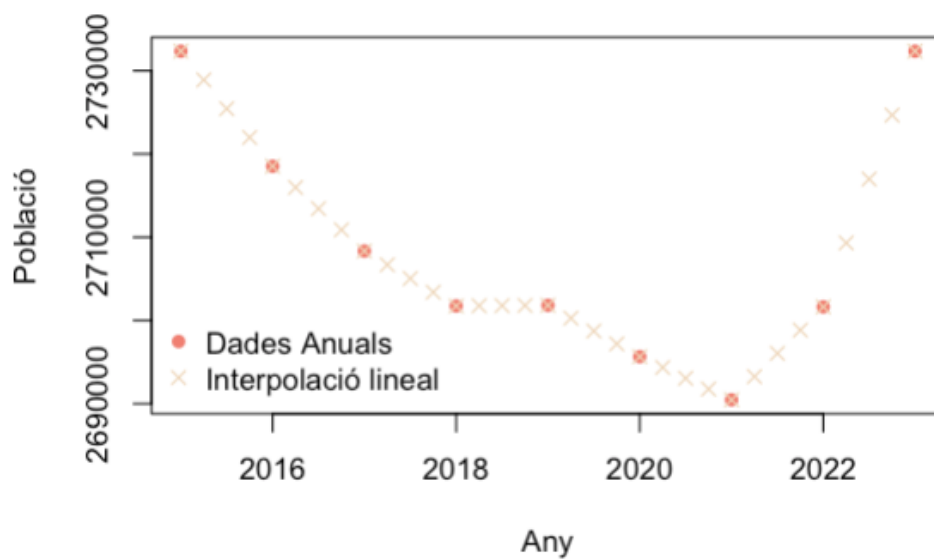
Font: Elaboració pròpia a partir de l'R i les dades de l'INE.

Gràfic 26. Interpolació lineal de les dades poblacionals d'Extremadura



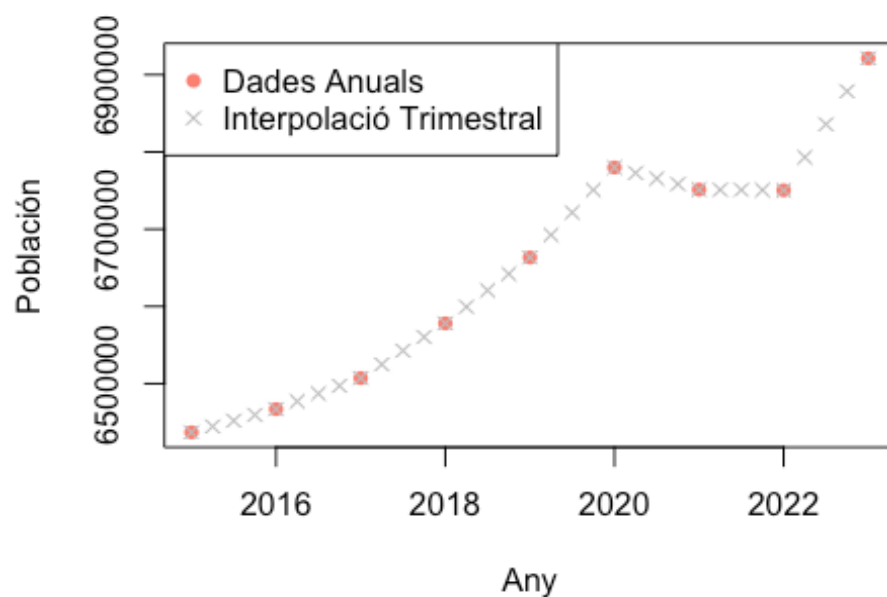
Font: Elaboració pròpia a partir de l'R i les dades de l'INE.

Gràfic 27. Interpolació lineal de les dades poblacionals de Galícia



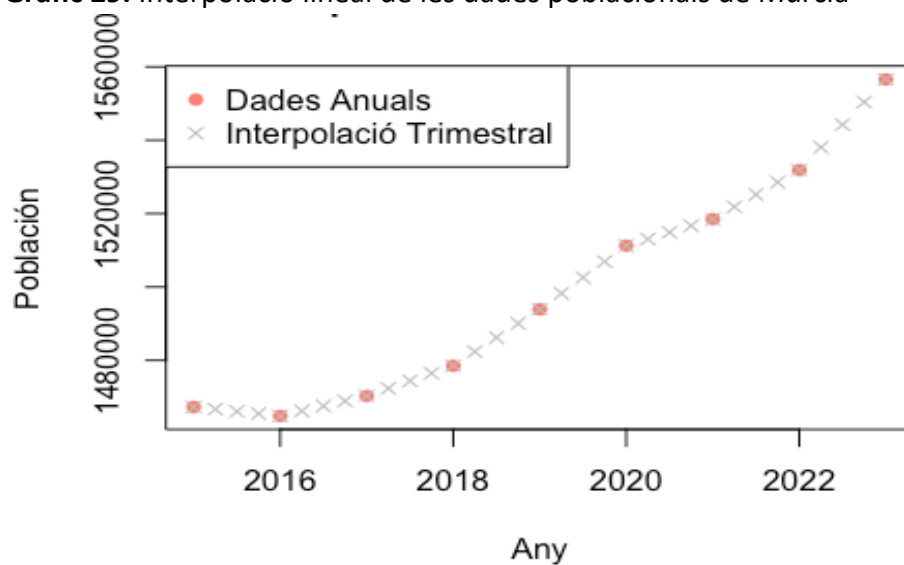
Font: Elaboració pròpia a partir de l'R i les dades de l'INE.

Gràfic 28. Interpolació lineal de les dades poblacionals de Madrid



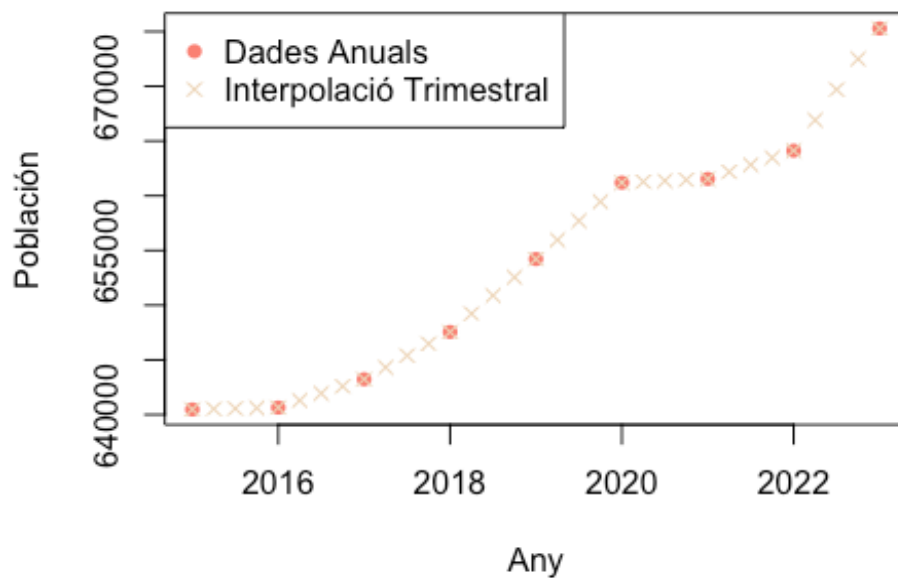
Font: Elaboració pròpia a partir de l'R i les dades de l'INE.

Gràfic 29. Interpolació lineal de les dades poblacionals de Múrcia



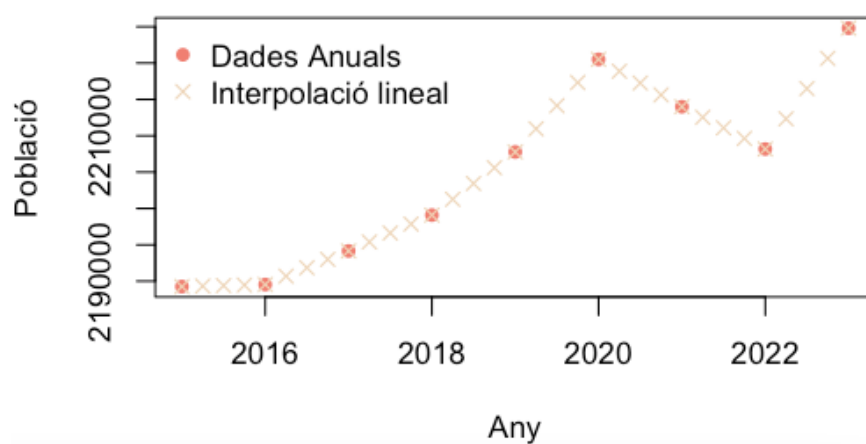
Font: Elaboració pròpia a partir de l'R i les dades de l'INE.

Gràfic 30. Interpolació lineal de les dades poblacionals de Navarra



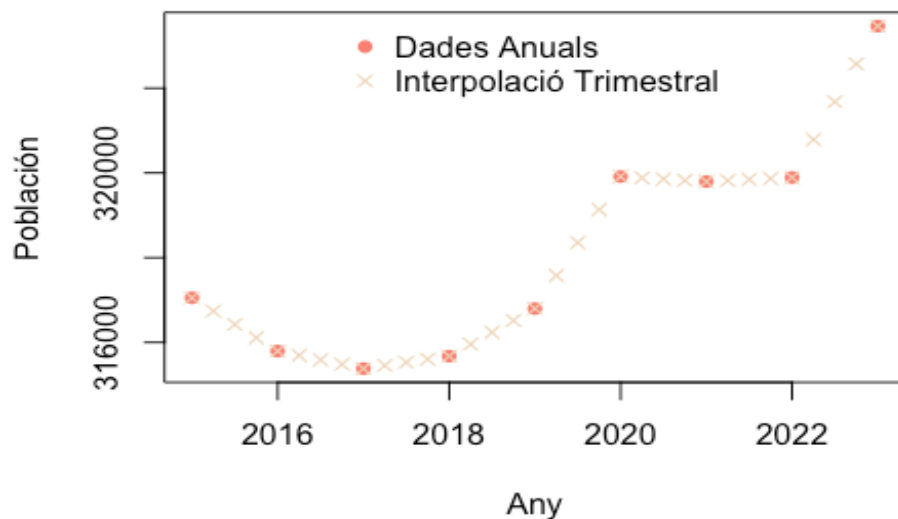
Font: Elaboració pròpia a partir de l'R i les dades de l'INE.

Gràfic 31. Interpolació lineal de les dades poblacionals del País Basc



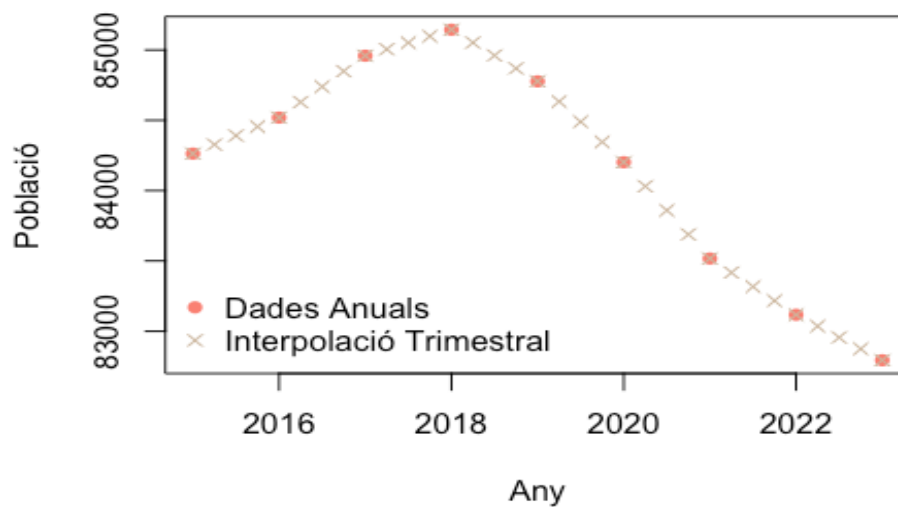
Font: Elaboració pròpia a partir de l'R i les dades de l'INE.

Gràfic 32. Interpolació lineal de les dades poblacionals de La Rioja



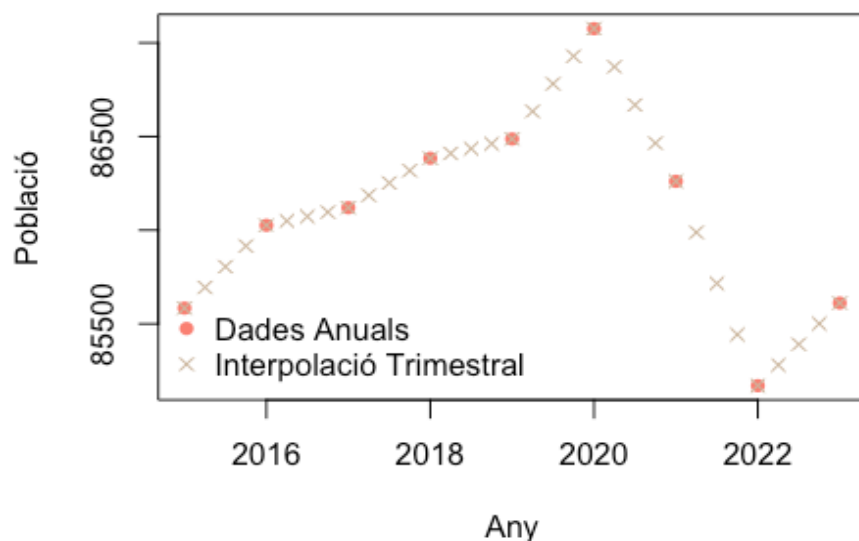
Font: Elaboració pròpia a partir de l'R i les dades de l'INE.

Gràfic 33. Interpolació lineal de les dades poblacionals de Ceuta



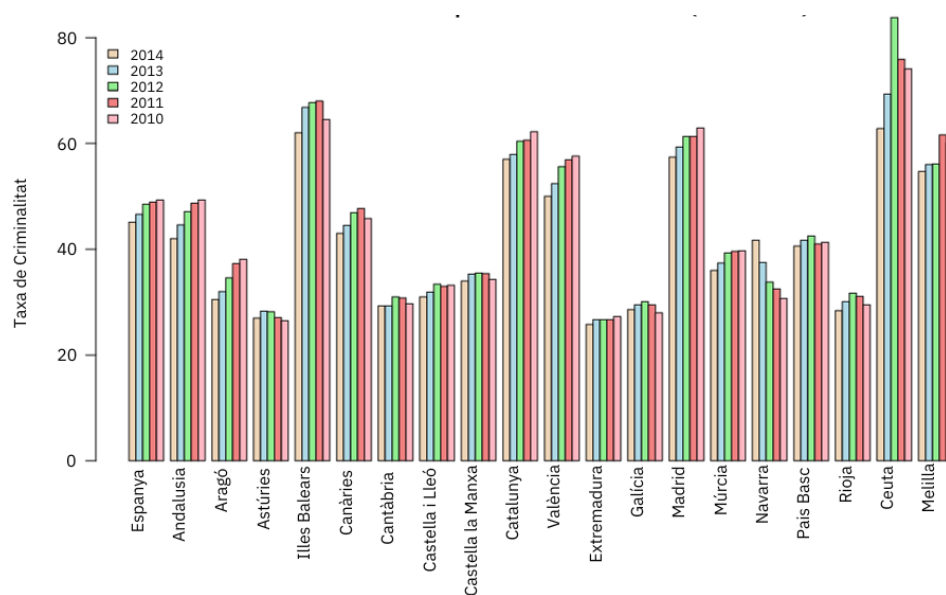
Font: Elaboració pròpia a partir de l'R i les dades de l'INE.

Gràfic 34. Interpolació lineal de les dades poblacionals de Melilla



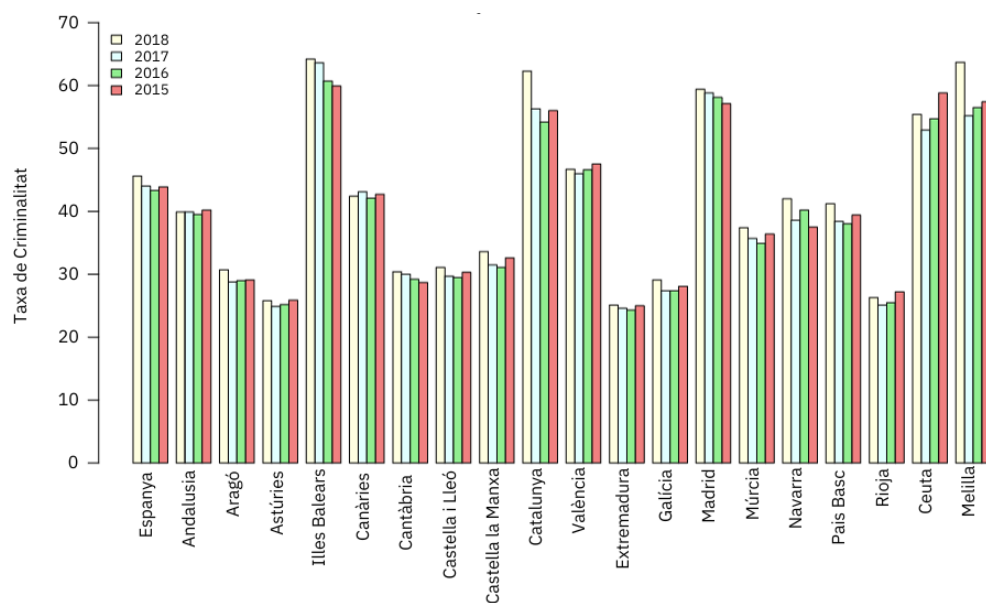
Font: Elaboració pròpia a partir de l'R i les dades de l'INE.

Gràfic 35. Taxa de Criminalitat per CCAA. Període 2010-2014.



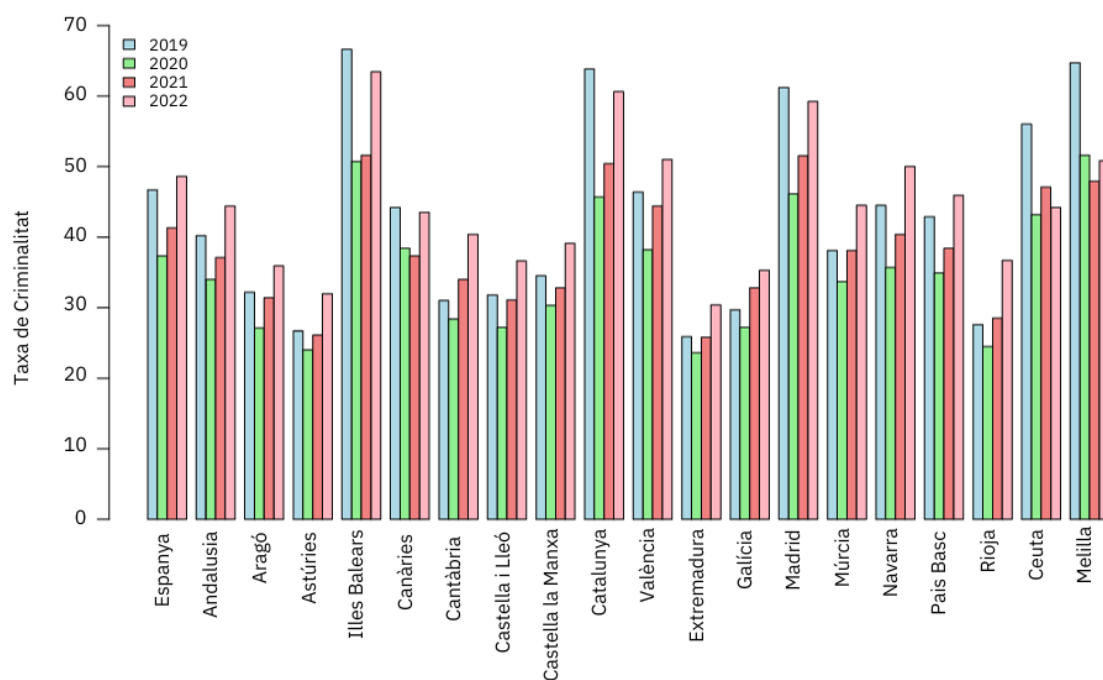
Font: Elaboració pròpia a partir de l'R i les dades del Ministeri d'Interior.

Gràfic 36. Taxa de Criminalitat per CCAA. Període 2015-2018.



Font: Elaboració pròpia a partir de l'R i les dades del Ministeri d'Interior.

Gràfic 37. Taxa de Criminalitat per CCAA. Període 2019-2022.



Font: Elaboració pròpia a partir de l'R i les dades del Ministeri d'Interior.

B. Taules complementàries

Figura 5. Divergència entre classificacions. 2016-2017.

	Variación % 2018/2017		Enero-marzo 2015	Enero-marzo 2016	Variación % 2016/2015
ESPAÑA		ESPAÑA			
I. CRIMINALIDAD CONVENCIONAL	-5,9	2.-HOMICIDIOS DOLOSOS Y ASESINATOS CONSUMADOS			
1. Homicidios dolosos y asesinatos consumados	-0,1	(EU)	66	72	9,1
2. Homicidios dolosos y asesinatos en grado tentativa	0,9				
3. Delitos graves y menos graves de lesiones y riña tumultuaria	17,4	1.-DELITOS Y FALTAS (EU)	496.876	487.643	-1,9
4. Secuestro	18,1				
5. Delitos contra la libertad sexual	22,7				
5.1.-Agresión sexual con penetración	17,5				
5.2.-Resto de delitos contra la libertad sexual	-1,8				
6. Robos con violencia e intimidación	0,6	3.-ROBO CON VIOLENCIA E INTIMIDACIÓN (EU)	15.727	15.559	-1,1
7. Robos con fuerza en domicilios, establecimientos y otras instalaciones	1,8				
7.1.-Robos con fuerza en domicilios	-1,0	4.-ROBOS CON FUERZA EN DOMICILIOS (EU)	28.916	31.154	7,7
8. Hurtos	-15,0	8.-HURTOS	168.971	169.129	0,1
9. Sustracciones de vehículos	9,0	5.-SUSTRACCIÓN VEHÍCULOS A MOTOR (EU)	10.351	10.177	-1,7
10. Tráfico de drogas		6.-TRÁFICO DE DROGAS (EU)	2.878	2.699	-6,2
11. Resto de criminalidad convencional	9	7.-DAÑOS	54.028	53.517	-0,9
II. CIBERCRIMINALIDAD (infracciones penales cometidas en/por medio ciber)					
12.-Estafas informáticas					
13.-Otros ciberdelitos					
III. TOTAL INFRACCIONES PENALES	4				

C. Codi en llenguatge R

Interpolació lineal i taxa de criminalitat trimestral

Lectura de les dades

ESpanya

```
library("readxl")
```

```
Historic_Espanya <- read_excel("Historic Espanya.xlsx", sheet = "Trimestres")
```

Necessitem tenir les dades trimestrals reals. Com en el Ministeri d'Interior ens facil·liten l'acumulat hem de restar els acumulats per tenir els trimestrals. Aquestes operacions s'han fet a l'Excel.

```
Historic_Catalunya <- read_excel("Historic Catalunya.xlsx", sheet = "Trimestres")
```

```
Historic_Balears <- read_excel("Historic Balears .xlsx", sheet = "Trimestres")
```

```
Historic_Asturies <- read_excel("Historic Asturias.xlsx", sheet = "Trimestres")
```

```
Historic_Arago <- read_excel("Historic Aragón.xlsx", sheet = "Trimestres")
```

```
Historic_Andalusia <- read_excel("Historic Andalusia.xlsx", sheet = "Trimestres")
```

```
Historic_Canaries <- read_excel("Historic Canàries.xlsx", sheet = "Trimestres")
```

```
Historic_Cantabria <- read_excel("Historic Cantabria.xlsx", sheet = "Trimestres")
```

```
Historic_Castella_i_Lleo <- read_excel("Historic Castella i Lleó.xlsx", sheet = "Trimestres")
```

```
Historic_Castella_la_Manxa <- read_excel("Historic Castella la Manxa.xlsx", sheet = "Trimestres")
```

```

Historic_Valencia <- read_excel("Historic Comunitat Valenciana.xlsx", sheet = "Trimestres")
Historic_Extremadura <- read_excel("Historic Extremadura.xlsx", sheet = "Trimestres")
Historic_Galicia <- read_excel("Historic Galicia.xlsx", sheet = "Trimestres")
Historic_Madrid <- read_excel("Historic Madrid.xlsx", sheet = "Trimestres")
Historic_Murcia <- read_excel("Historic Murcia.xlsx", sheet = "Trimestres")
Historic_Navarra <- read_excel("Historic Navarra.xlsx", sheet = "Trimestres")
Historic_Pais_Basc <- read_excel("Historic Pais Basc.xlsx", sheet = "Trimestres")
Historic_Rioja <- read_excel("Historic Rioja.xlsx", sheet = "Trimestres")
Historic_Ceuta <- read_excel("Historic Ceuta.xlsx", sheet = "Trimestres")
Historic_Melilla <- read_excel("Historic Melilla.xlsx", sheet = "Trimestres")

# Utilitzem la funció gather per convertir les columnes en files
datos_Espanya_long <- Historic_Espanya[1,] %>%
  gather(key = "Trimestre_Any", value = "Crims") %>%
  mutate(Any = as.integer(substr(Trimestre_Any, 1, 4)),
         Trimestre = as.integer(substr(Trimestre_Any, 6, 6)))

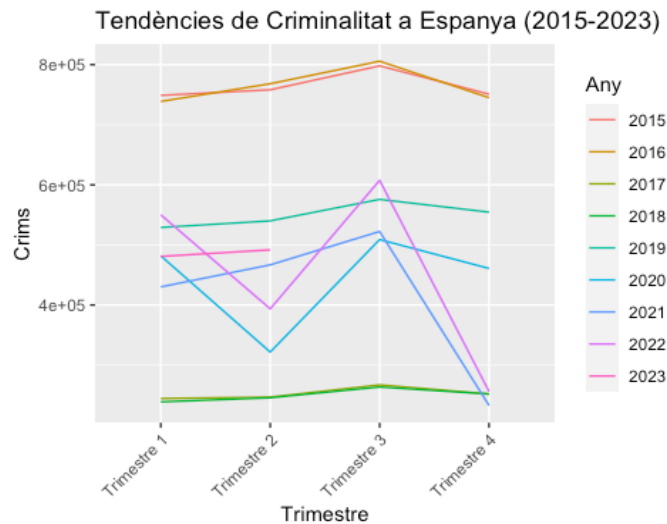
datos_transformados <- datos_Espanya_long %>%
  select(Any, Trimestre, Crims)

Crims_Espanya <- datos_transformados %>%
  arrange(Any)

# Modificar la variable Trimestre
Crims_Espanya$Trimestre <- factor(Crims_Espanya$Trimestre, levels = 1:4,
                                labels = c("Trimestre 1", "Trimestre 2", "Trimestre 3", "Trimestre 4"))

# Crear el gràfic de tendències
ggplot(Crims_Espanya, aes(x = Trimestre, y = Crims, group = Any, color = factor(Any))) +
  geom_line() +
  labs(title = "Tendències de Criminalitat a Espanya (2015-2023)",
       x = "Trimestre",
       y = "Crims") +
  scale_color_discrete(name = "Any") +
  theme(axis.text.x = element_text(angle = 45, hjust = 1)) # Rotar etiquetas del eje x para mejorar la legibilidad

```



Taxa de Criminalitat Trimestral

Espanya

```
poblacio_Espanya <- read_excel("Dades Trimestrals demografia Espanya.xlsx", range = "B7:F16", col_
types = c("numeric", "numeric", "numeric", "numeric", "numeric"))
poblacio_Espanya <- poblacio_Espanya %>% arrange(...1)
poblacio_Espanya[9,4:5] <- NA
poblacio_Espanya <- poblacio_Espanya %>%
  pivot_longer(cols = starts_with("T"),
               names_to = "Trimestre",
               values_to = "Poblacion") %>%
  mutate(Trimestre = gsub("T", "Trimestre", Trimestre, fixed = TRUE)) %>%
  arrange(...1, Trimestre)

poblacio_Espanya <- head(poblacio_Espanya, -2)

# TAXA DE CRIMINALITAT
## Tasa de Criminalidad Trimestral = (Número de Delitos en ese Trimestre / Población en ese Trime
stre) * Factor de Escala

library(dplyr)

poblacio_Espanya <- poblacio_Espanya %>%
  rename(Any = ...1, Trimestre = Trimestre, Poblacion = Poblacion)

# Unir las tablas por año y trimestre
merged_data <- inner_join(poblacio_Espanya, Crims_Espanya, by = c("Any", "Trimestre"))

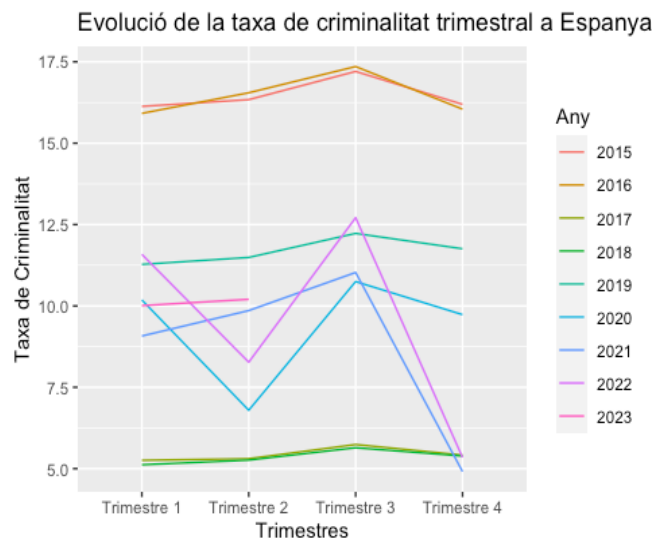
# Calcular la tasa de criminalidad
taxa_criminalitat_Espanya <- merged_data %>%
  mutate(Tasa_Criminalidad = (Crims / Poblacion) * 1000)

# REPRESENTACIÓ GRÀFICA
library(ggplot2)
# Gràfic de línia amb ggplot2
ggplot(taxa_criminalitat_Espanya, aes(x = Trimestre, y = Tasa_Criminalidad, group = Any, color = fact
```

```

or(Any))) +
  geom_line() +
  labs(title = "Evolució de la taxa de criminalitat trimestral a Espanya (2015-2023)",
        x = "Trimestres",
        y = "Taxa de Criminalitat") +
  scale_color_discrete(name = "Any")

```



Catalunya

```

dades_anuals_Cat <- data.frame(Any = c(2015, 2016, 2017, 2018, 2019, 2020, 2021, 2022, 2023), Poblacio = c(7430638, 7455671, 7504764, 7553681, 7629889, 7732756, 7749896, 7770689, 7899056))

```

```

trimestres <- seq(dades_anuals_Cat$Any[1], dades_anuals_Cat$Any[length(dades_anuals_Cat$Any)], by = 0.25)

```

Interpolador linealment los datos trimestrales

```

interpolacion <- approx(x = dades_anuals_Cat$Any, y = dades_anuals_Cat$Poblacio, xout = trimestres, method = "linear")

```

```

dades_trimestrals_Cat <- data.frame(Any = interpolacion$x, Poblacio = interpolacion$y)

```

Utilitzem la funció gather per convertir les columnes en files

```

datos_Catalunya_long <- Historic_Catalunya[1,] %>%
  gather(key = "Trimestre_Any", value = "Crims") %>%
  mutate(Any = as.integer(substr(Trimestre_Any, 1, 4)),
         Trimestre = as.integer(substr(Trimestre_Any, 6, 6)))

```

```

datos_transformados <- datos_Catalunya_long %>%
  select(Any, Trimestre, Crims)

```

```

Crims_Catalunya <- datos_transformados %>%
  arrange(Any)

```

Modificar la variable Trimestre

```

Crims_Catalunya$Trimestre <- factor(Crims_Catalunya$Trimestre, levels = 1:4, labels = c("Trimestre 1", "Trimestre 2", "Trimestre 3", "Trimestre 4"))

```

```

# Dades poblacionals:
# Convertir la columna "Any" a factor
dades_trimestrals_Cat$Any <- as.factor(dades_trimestrals_Cat$Any)

# Crear la columna "Trimestres"
trimestres <- rep(c("Trimestre 1", "Trimestre 2", "Trimestre 3", "Trimestre 4"), length.out = nrow(dades_trimestrals_Cat))

# Afegir la nova columna "Trimestres"
dades_trimestrals_Cat$Trimestres <- rep(trimestres, each = 1)
dades_trimestrals_Cat$Any <- c(rep(2015:2022, each=4), 2023)

# Reorganizar las columnas
dades_trimestrals_Cat <- dades_trimestrals_Cat[, c("Any", "Trimestres", "Poblacio")]
# Cambiar el nombre de la columna
colnames(dades_trimestrals_Cat)[colnames(dades_trimestrals_Cat) == "Trimestres"] <- "Trimestre"

# TAXA DE CRIMINALITAT
## Tasa de Criminalidad Trimestral = (Número de Delitos en ese Trimestre / Población en ese Trimestre) * Factor de Escala (1.000)
merged_data <- inner_join(dades_trimestrals_Cat, Crims_Catalunya, by = c("Any", "Trimestre"))

# Calcular la tasa de criminalidad
taxa_criminalitat_Catalunya <- merged_data %>%
  mutate(Tasa_Criminalidad = (Crims / Poblacio) * 1000)

# REPRESENTACIÓ GRÀFICA:
ggplot(taxa_criminalitat_Catalunya[1:32,], aes(x = Trimestre, y = Tasa_Criminalidad, group = Any, color = factor(Any))) +
  geom_line() +
  labs(title = "Evolució de la taxa de criminalitat trimestral a Catalunya (2015-2022)",
    x = "Trimestres",
    y = "Taxa de Criminalitat") +
  scale_color_discrete(name = "Any") +
  ylim(c(0, 25))

```

Es realitzen les mateixes passes que s'han realitzat a Catalunya per a la resta de regions.

Models de recomptes

Explicació variables del model:

- **Taxa d'envelliment:** índex d'envelliment en %.
- **Estrangers:** població estrangera que viu a Espanya. Unitats: persones estandarditzades per la població.
- **Turisme:** número de turistes segons la CCAA de destí principal, queden exclosos els turistes nacionals. Només es contemplen els estrangers. Estandarditzats per la població
- **PIB pc:** PIB per càpita expressat en €.

- **Renda mitjana llars:** Renda mitjana de les llars en €. S'inclou l'arrendament imputat.
- **Risc pobresa:** Taxa de risc de pobresa per CCAA en %. Usa la mitjana dels ingressos ajustats per unitat de consum com a punt de referència i defineix el llindar de pobresa com el 60% d'aquesta mediana.
- **Aturats:** aturats per CCAA en milers de persones estandarditzats per la població.
- **Inferior a 2ª etapa E. Secundària:** percentatge de persones amb nacionalitat espanyola entre 25-64 anys que tenen el nivell d'estudis inferior a la 2a etapa d'Educació Secundària.
- **2ª etapa E. Secundària:** percentatge de persones amb nacionalitat espanyola entre 25-64 anys que tenen el nivell d'estudis corresponent a la 2a etapa d'Educació Secundària.
- **E. Superior:** percentatge de persones amb nacionalitat espanyola entre 25-64 anys que tenen el nivell d'estudis corresponent l'Educació Superior.
- **Taxa abandonament prematur escolar:** taxa d'abandonament educatiu aviat de la població entre 18 a 24 anys per CCAA.
- **Divorcis:** número de divorcis per mil habitants, taxa bruta de divorcis per CCAA.
- **Població:** població potencial a cometre un delictes, es troba el número d'habitants de 16 anys i més.

Descriptiva

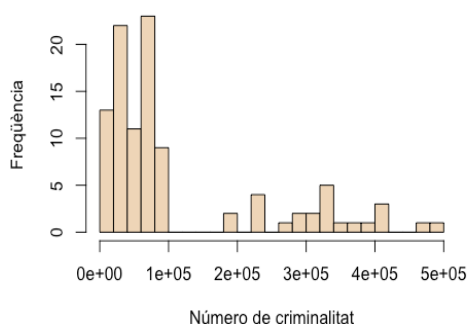
Es procedeix a fer una descriptiva de les dades incloent una descriptiva bivariant de la resposta amb la resta de variables del model.

```
summary(taxa_factors)
```

```
# Histograma de la variable resposta
```

```
hist(taxa_factors$Resta_delictes,breaks=20, col = "bisque2", main= "Histograma de la taxa de criminalitat convencional total", xlab= "Número de criminalitat", ylab = "Freqüència") # variable resposta
```

Histograma de la taxa de criminalitat convencional t

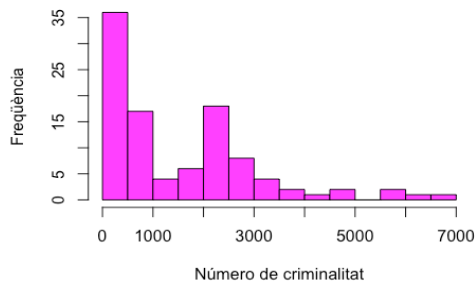


```
# Histograma de la variable resposta
```



```
hist(taxa_factors$Delictes_contra_persona,breaks=20, col = "magenta", main= "Histograma de la taxa de criminalitat convencional total", xlab= "Número de criminalitat", ylab = "Freqüència") # variable resposta
```

Histograma de la taxa de criminalitat convencional t



La variable resposta presenta una distribució assimètrica, característica de les dades de comptatges.

Descriptiva bivariant: Taxa de Criminalitat per CCAA

Boxplot:

```
par(las = 2) # Rotar los nombres de las CCAA en el eje x
```

```
boxplot(taxa_factors$Resta_delictes/taxa_factors$Pobl16 ~ taxa_factors$CCAA,
        main = "Delictes menors per CCAA",
        xlab = "", ylab = "Taxa de criminalitat",
        col = rainbow(length(unique(taxa_factors$CCAA))))
```

```
par(las = 1)
```

Descriptiva bivariant: Taxa de Criminalitat per CCAA

Boxplot:

```
par(las = 2)
```

```
boxplot(taxa_factors$Delictes_contra_persona/taxa_factors$Pobl16 ~ taxa_factors$CCAA,
        main = "Boxplot de Criminalidad por CCAA",
        xlab = "", ylab = "Criminalidad",
        col = rainbow(length(unique(taxa_factors$CCAA))))
```

```
par(las = 1)
```

install.packages("corrplot")

```
library(corrplot)
```

Crear una matriz de correlación

```
cor_matrix <- cor(taxa_factors[, c("Taxa_envelliment", "Estrangers", "Turisme", "PIB_pc", "Renda_mitjana_
    llars", "Risc_pobresa", "Aturats", "Divorcis", "Inferior_2a_etapa_E.Secundaria", "Taxa_abandonament_a
    viat_escolar"])]
```

Visualizar la matriz de correlación

```
corrplot::corrplot(cor_matrix, method = "color")
```

```
variables_numericas <- c("Estrangers", "Turisme", "PIB_pc", "Renda_mitjana_llars", "Risc_pobresa", "Aturats", "Inferior_2a_etapa_E.Secundaria", "Segona_etapa_E.Secundaria", "E.Superior", "Taxa_abandonament_aviat_escolar", "Divorcis", "Pobl16")
```

Variable objetivo: Resta_delictes

```
for (variable in variables_numericas) {
```

```
  # Crear el gráfico bivalente
```

```

plot(taxa_factors$Resta_delictes, taxa_factors[[variable]], main = paste("Resta_delictes vs.", variable),
     xlab = "Resta_delictes", ylab = variable, pch = 16, col = "bisque4")
}

# Variable objetivo: va_indicadora
for (variable in variables_numericas) {
  boxplot(taxa_factors[[variable]] ~ taxa_factors$va_indicadora, data = taxa_factors,
          main = paste("Variable indicadora vs.", variable),
          xlab = "Variable Indicadora de la pandèmia", ylab = variable,
          col = c("lightblue", "lightpink"))
}

# Variable objetivo: Delictes_contra_persona
for (variable in variables_numericas) {
  # Crear el gráfico bivalente
  plot(taxa_factors$Delictes_contra_persona, taxa_factors[[variable]], main = paste("Delictes_contra_persona vs.", variable),
       xlab = "Delictes_contra_persona", ylab = variable, pch = 16, col = "bisque1")
}

boxplot(taxa_factors$Resta_delictes ~ taxa_factors$va_indicadora, data = taxa_factors,
        main = paste("Variable indicadora vs. Resta de delictes"),
        xlab = "Variable Indicadora de la pandèmia", ylab = variable,
        col = c("lightpink", "bisque2"))

boxplot(taxa_factors$Delictes_contra_persona ~ taxa_factors$va_indicadora, data = taxa_factors,
        main = paste("Variable indicadora vs. Delictes contra persones"),
        xlab = "Variable Indicadora de la pandèmia", ylab = variable,
        col = c("lightblue", "lightgreen"))

```

Model 1 - Resta de criminalitat

Model de Poisson

```

df <- data.frame(taxa_factors)
df$Anys <- NULL

```

Ahora asturias_df contiene solo las filas donde la Provincia es "Asturias"

```

df$CCAA <- NULL
df$Delictes_contra_persona <- NULL

```

Realizar la selección de variables hacia atrás con stepAIC

```

modelo_poisson1 <- stepAIC(model1_full, direction = "backward", k = log(nrow(df)), trace = 0)

```

```

summary(modelo_poisson1)

```

```

## Call:

```

```

## glm(formula = Resta_delictes ~ offset(log(Pobl16)) + Taxa_envelliment +
##   Estrangers + Turisme + PIB_pc + Renda_mitjana_llars + Risc_pobresa +
##   Aturats + Inferior_2a_etapa_E.Secundaria + Segona_etapa_E.Secundaria +
##   E.Superior + Taxa_abandonament_aviat_escolar + Divorcis +
##   va_indicadora + Turisme * va_indicadora, family = poisson,
##   data = df)
##

```

```

## Deviance Residuals:
##   Min     1Q   Median     3Q      Max
## -69.133 -11.953  1.328  13.982  54.962
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)      1.815e+04  7.285e+02  24.91 <2e-16 ***
## Taxa_envelliment    -2.677e-03  2.375e-05 -112.70 <2e-16 ***
## Estrangers         1.946e+00  1.770e-02  109.94 <2e-16 ***
## Turisme           2.758e-01  4.751e-03   58.04 <2e-16 ***
## PIB_pc            1.727e-05  2.539e-07   68.03 <2e-16 ***
## Renda_mitjana_llars -1.073e-05  2.917e-07  -36.77 <2e-16 ***
## Risc_pobresa       -1.136e-02  1.236e-04  -91.94 <2e-16 ***
## Aturats           7.318e+03  1.086e+02   67.39 <2e-16 ***
## Inferior_2a_etapa_E.Secundaria -1.815e+02  7.285e+00  -24.92 <2e-16 ***
## Segona_etapa_E.Secundaria  -1.815e+02  7.285e+00  -24.92 <2e-16 ***
## E.Superior        -1.815e+02  7.285e+00  -24.92 <2e-16 ***
## Taxa_abandonament_aviat_escolar 1.916e-03  1.848e-04  10.37 <2e-16 ***
## Divorcis          1.302e-01  2.384e-03   54.60 <2e-16 ***
## va_indicadora     -6.132e-02  1.403e-03  -43.71 <2e-16 ***
## Turisme:va_indicadora -3.645e-01  1.527e-02  -23.87 <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for poisson family taken to be 1)
##
##   Null deviance: 898134  on 101  degrees of freedom
## Residual deviance: 63304  on 87  degrees of freedom
## AIC: 64655
##
## Number of Fisher Scoring iterations: 4

```

D'entrada surten totes les variables significatives, però això no vol dir que el model estigui bé, hem de mirar la sobredispersió, que en el cas d'haver-hi, s'estarien infravalorant els errors estàndards, això explicaria perquè surten tants coeficients significatius.

```

# Estimar paràmetre de sobredispersió
X2P <- sum(resid(modelo_poisson1,type="pearson")^2)
phi <- X2P/modelo_poisson1$df.residual
paste0("El paràmetre de sobredispersió estimat és ", round(phi,3))

## [1] "El paràmetre de sobredispersió estimat és 714.374"

# Mirem test de sobredispersió: Trafo = 1 & Trafo = 2:

dispersiontest(modelo_poisson1, trafo = 1) #  $V(Y) = (1+\alpha)*\mu \rightarrow Y$  Quasi Poisson
## Overdispersion test
##
## data: modelo_poisson1
## z = 6.5418, p-value = 3.04e-11
## alternative hypothesis: true alpha is greater than 0
## sample estimates:
##  alpha
## 608.318

```

```
dispersiontest(modelo_poisson1, trafo = 2) #  $V(Y) = \mu + \alpha * \mu^2 \rightarrow Y \text{ Bin Neg}$ 
```

```
## Overdispersion test
##
## data: modelo_poisson1
## z = 5.8188, p-value = 2.964e-09
## alternative hypothesis: true alpha is greater than 0
## sample estimates:
##    alpha
## 0.003319993
```

Model Binomial negatiu

```
mod.nb01 <- glm.nb(Resta_delictes ~ offset(log(Pobl16)) + Taxa_envelliment + Estrangers + Turisme + PIB_pc + Renda_mitjana_llars + Risc_pobresa + Aturats + Inferior_2a_etapa_E.Secundaria + Segona_etapa_E.Secundaria + E.Superior + Taxa_abandonament_aviat_escolar + Divorcis + va_indicadora, data=df)
```

```
mod.nb001 <- glm(Resta_delictes ~ offset(log(Pobl16)) + Taxa_envelliment + Estrangers + Turisme + PIB_pc + Renda_mitjana_llars + Risc_pobresa + Aturats + Inferior_2a_etapa_E.Secundaria + Segona_etapa_E.Secundaria + E.Superior + Taxa_abandonament_aviat_escolar + Divorcis + va_indicadora, data=df, family = negative.binomial(mod.nb01$theta))
```

```
summary(mod.nb01)
```

```
##
## Call:
## glm.nb(formula = Resta_delictes ~ offset(log(Pobl16)) + Taxa_envelliment +
## Estrangers + Turisme + PIB_pc + Renda_mitjana_llars + Risc_pobresa +
## Aturats + Inferior_2a_etapa_E.Secundaria + Segona_etapa_E.Secundaria +
## E.Superior + Taxa_abandonament_aviat_escolar + Divorcis +
## va_indicadora, data = df, init.theta = 105.3976149, link = log)
##
## Deviance Residuals:
##    Min       1Q   Median       3Q      Max
## -3.06324 -0.54983 -0.05131  0.39863  2.36121
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)    1.975e+04  2.126e+04  0.929 0.352839
## Taxa_envelliment    -1.993e-03  5.492e-04 -3.629 0.000285 ***
## Estrangers        2.329e+00  4.244e-01  5.487 4.08e-08 ***
## Turisme          3.696e-01  1.508e-01  2.451 0.014248 *
## PIB_pc           1.053e-05  6.738e-06  1.562 0.118262
## Renda_mitjana_llars    4.763e-06  8.014e-06  0.594 0.552246
## Risc_pobresa      -1.376e-02  3.123e-03 -4.405 1.06e-05 ***
## Aturats          5.283e+03  2.864e+03  1.845 0.065084 .
## Inferior_2a_etapa_E.Secundaria -1.975e+02  2.126e+02 -0.929 0.352770
## Segona_etapa_E.Secundaria    -1.976e+02  2.126e+02 -0.929 0.352717
## E.Superior        -1.975e+02  2.126e+02 -0.929 0.352754
## Taxa_abandonament_aviat_escolar -7.922e-04  4.898e-03 -0.162 0.871509
## Divorcis          1.551e-01  6.444e-02  2.408 0.016062 *
## va_indicadora      -8.308e-02  3.134e-02 -2.651 0.008034 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for Negative Binomial(105.3976) family taken to be 1)
```

```
##
## Null deviance: 965.06 on 101 degrees of freedom
## Residual deviance: 102.26 on 88 degrees of freedom
## AIC: 2112.2
##
## Number of Fisher Scoring iterations: 1
##
##
##      Theta: 105.4
##      Std. Err.: 14.8
##
## 2 x log-likelihood: -2082.248
```

El resultat de corregir la sobredispersió que presentava el model de Poisson es que varies variables deixin de ser significatives perquè amb sobredispersió s'infravaloren els errors estàndards, passem de tenir totes les variables explicatives significatives a només tenir 7 significatives.

```
mod2.nb01 <- glm.nb(Resta_delictes ~ offset(log(Pobl16)) + Taxa_envelliment + Estrangers + Turisme + PIB_pc + Renda_mitjana_llars + Risc_pobresa + Aturats + Segona_etapa_E.Secundaria + E.Superior + Taxa_abandonament_aviat_escolar + Divorcis + va_indicadora, data=df)
```

```
mod2.nb001 <- glm(Resta_delictes ~ offset(log(Pobl16)) + Taxa_envelliment + Estrangers + Turisme + PIB_pc + Renda_mitjana_llars + Risc_pobresa + Aturats + Segona_etapa_E.Secundaria + E.Superior + Taxa_abandonament_aviat_escolar + Divorcis + va_indicadora, data=df, family = negative.binomial(mod2.nb01$theta))
```

```
summary(mod2.nb001)
```

```
anova(mod.nb01, mod2.nb01, test = "Chisq")
```

```
## Likelihood ratio tests of Negative Binomial Models
```

```
##
```

```
## Response: Resta_delictes
```

```
##
```

```
Model
```

```
## 1 offset(log(Pobl16)) + Taxa_envelliment + Estrangers + Turisme + PIB_pc + Renda_mitjana_llars + Risc_pobresa + Aturats + Segona_etapa_E.Secundaria + E.Superior + Taxa_abandonament_aviat_escolar + Divorcis + va_indicadora
```

```
## 2 offset(log(Pobl16)) + Taxa_envelliment + Estrangers + Turisme + PIB_pc + Renda_mitjana_llars + Risc_pobresa + Aturats + Inferior_2a_etapa_E.Secundaria + Segona_etapa_E.Secundaria + E.Superior + Taxa_abandonament_aviat_escolar + Divorcis + va_indicadora
```

```
## theta Resid. df 2 x log-lik. Test df LR stat. Pr(Chi)
```

```
## 1 104.5180 89 -2083.105
```

```
## 2 105.3976 88 -2082.248 1 vs 2 1 0.8572187 0.3545182
```

No hay suficiente evidencia para decir que la inclusión de la variable **Inferior_2a_etapa_E.Secundaria** mejora significativamente la bondad de ajuste del modelo. Por lo tanto, se decide optar por el modelo que no incluye la variable **Inferior_2a_etapa_E.Secundaria**.

Si miramos los BIC, el modelo sin la variable **Inferior_2a_etapa_E.Secundaria** también presenta el mínimo:

```
BIC(mod.nb01, mod2.nb01)
```

```
##      df      BIC
## mod.nb01 15 2151.622
## mod2.nb01 14 2147.854

mod3.nb01 <- glm.nb(Resta_delictes ~ offset(log(Pobl16)) + Taxa_envelliment + Estrangers + Turisme + PIB_pc + Risc_pobresa + Aturats + Segona_etapa_E.Secundaria + E.Superior + Taxa_abandonament_aviat_escolar + Divorcis + va_indicadora, data=df)

mod3.nb001 <- glm(Resta_delictes ~ offset(log(Pobl16)) + Taxa_envelliment + Estrangers + Turisme + PIB_pc + Risc_pobresa + Aturats + Segona_etapa_E.Secundaria + E.Superior + Taxa_abandonament_aviat_escolar + Divorcis + va_indicadora, data=df, family = negative.binomial(mod3.nb01$theta))

anova(mod2.nb01, mod3.nb01, test = "Chisq")

## Likelihood ratio tests of Negative Binomial Models
##
## Response: Resta_delictes
##
Model
## 1      offset(log(Pobl16)) + Taxa_envelliment + Estrangers + Turisme + PIB_pc + Risc_pobresa +
Aturats + Segona_etapa_E.Secundaria + E.Superior + Taxa_abandonament_aviat_escolar + Divorcis + va_i
ndicadora
## 2 offset(log(Pobl16)) + Taxa_envelliment + Estrangers + Turisme + PIB_pc + Renda_mitjana_llars + Risc
_pobresa + Aturats + Segona_etapa_E.Secundaria + E.Superior + Taxa_abandonament_aviat_escolar + Div
orcis + va_indicadora
##      theta Resid. df    2 x log-lik. Test    df LR stat. Pr(Chi)
## 1 104.2811     90    -2083.332
## 2 104.5180     89    -2083.105 1 vs 2    1 0.2275124 0.6333745

BIC(mod2.nb01, mod3.nb01)

##      df      BIC
## mod2.nb01 14 2147.854
## mod3.nb01 13 2143.457
```

El resultado del test indica que no hay evidencia suficiente para rechazar la hipótesis nula de que los dos modelos son estadísticamente equivalentes.

En este caso, podrías considerar prescindir de la variable: **Renda_mitjana_llars** ya que la adición de estas variables al modelo no mejora significativamente su bondad de ajuste, al menos según la prueba de razón de verosimilitud.

```
mod4.nb01 <- glm.nb(Resta_delictes ~ offset(log(Pobl16)) + Taxa_envelliment + Estrangers + Turisme + PIB_pc + Risc_pobresa + Aturats + Segona_etapa_E.Secundaria + Taxa_abandonament_aviat_escolar + Divorcis + va_indicadora, data=df)

mod4.nb001 <- glm(Resta_delictes ~ offset(log(Pobl16)) + Taxa_envelliment + Estrangers + Turisme + PIB_pc + Risc_pobresa + Aturats + Segona_etapa_E.Secundaria + Taxa_abandonament_aviat_escolar + Divorcis + va_indicadora, data=df, family = negative.binomial(mod4.nb01$theta))

anova(mod3.nb01, mod4.nb01, test = "Chisq")

## Likelihood ratio tests of Negative Binomial Models
##
## Response: Resta_delictes
```

```
##
Model
## 1      offset(log(Pobl16)) + Taxa_envelliment + Estrangers + Turisme + PIB_pc + Risc_pobresa + Aturats + Segona_etapa_E.Secundaria + Taxa_abandonament_aviat_escolar + Divorcis + va_indicadora
## 2 offset(log(Pobl16)) + Taxa_envelliment + Estrangers + Turisme + PIB_pc + Risc_pobresa + Aturats + Segona_etapa_E.Secundaria + E.Superior + Taxa_abandonament_aviat_escolar + Divorcis + va_indicadora
##      theta Resid. df    2 x log-lik. Test    df LR stat. Pr(Chi)
## 1 102.9418    91    -2084.652
## 2 104.2811    90    -2083.332 1 vs 2    1 1.319637 0.2506572

BIC(mod3.nb01, mod4.nb01)

##      df    BIC
## mod3.nb01 13 2143.457
## mod4.nb01 12 2140.152
```

Podem prescindir de la interacció: **E.Superior**

```
mod5.nb01 <- glm.nb(Resta_delictes ~ offset(log(Pobl16)) + Taxa_envelliment + Estrangers + Turisme + Renda_mitjana_llars + Risc_pobresa + Inferior_2a_etapa_E.Secundaria + Segona_etapa_E.Secundaria + Divorcis + va_indicadora, data=df)

mod5.nb001 <- glm(Resta_delictes ~ offset(log(Pobl16)) + Taxa_envelliment + Estrangers + Turisme + Renda_mitjana_llars + Risc_pobresa + Inferior_2a_etapa_E.Secundaria + Segona_etapa_E.Secundaria + Divorcis + va_indicadora, data=df, family = negative.binomial(mod5.nb01$theta))

anova(mod4.nb01, mod5.nb01, test = "Chisq")

## Likelihood ratio tests of Negative Binomial Models
##
## Response: Resta_delictes
##
Model
## 1 offset(log(Pobl16)) + Taxa_envelliment + Estrangers + Turisme + Renda_mitjana_llars + Risc_pobresa + Inferior_2a_etapa_E.Secundaria + Segona_etapa_E.Secundaria + Divorcis + va_indicadora
## 2 offset(log(Pobl16)) + Taxa_envelliment + Estrangers + Turisme + PIB_pc + Risc_pobresa + Aturats + Segona_etapa_E.Secundaria + Taxa_abandonament_aviat_escolar + Divorcis + va_indicadora
##      theta Resid. df    2 x log-lik. Test    df LR stat. Pr(Chi)
## 1 99.44781    92    -2088.154
## 2 102.94185    91    -2084.652 1 vs 2    1 3.501847 0.06130041

BIC(mod5.nb01, mod4.nb01)

##      df    BIC
## mod5.nb01 11 2139.029
## mod4.nb01 12 2140.152
```

Podemos prescindir de **Aturats**.

```
mod6.nb01 <- glm.nb(Resta_delictes ~ offset(log(Pobl16)) + Taxa_envelliment + Estrangers + Turisme + Renda_mitjana_llars + Risc_pobresa + Segona_etapa_E.Secundaria + Divorcis + va_indicadora, data=df)

mod6.nb001 <- glm(Resta_delictes ~ offset(log(Pobl16)) + Taxa_envelliment + Estrangers + Turisme + Renda_mitjana_llars + Risc_pobresa + Segona_etapa_E.Secundaria + Divorcis + va_indicadora, data=df, family = negative.binomial(mod6.nb01$theta))

anova(mod6.nb001, mod5.nb001, test = "Chisq")
```

```
## Analysis of Deviance Table
##
## Model 1: Resta_delictes ~ offset(log(Pobl16)) + Taxa_envelliment + Estrangers +
##   Turisme + Renda_mitjana_llars + Risc_pobresa + Segona_etapa_E.Secundaria +
##   Divorcis + va_indicadora
## Model 2: Resta_delictes ~ offset(log(Pobl16)) + Taxa_envelliment + Estrangers +
##   Turisme + Renda_mitjana_llars + Risc_pobresa + Inferior_2a_etapa_E.Secundaria +
##   Segona_etapa_E.Secundaria + Divorcis + va_indicadora
##   Resid. Df Resid. Dev Df   Deviance Pr(>Chi)
## 1     93   102.24
## 2     92   102.24 1 -0.00062873
```

```
BIC(mod6.nb001, mod5.nb001)
```

```
##      df    BIC
## mod6.nb001 9 2130.067
## mod5.nb001 10 2134.404
```

La variable **Inferior_2a_etapa_E.Secundaria** té gran influència. Pel criteri del BIC, podem no incloure-la.

```
mod7.nb01 <- glm.nb(Resta_delictes ~ offset(log(Pobl16)) + Taxa_envelliment + Estrangers + Turisme + Risc_pobresa + Segona_etapa_E.Secundaria + Divorcis + va_indicadora, data=df)
```

```
mod7.nb001 <- glm(Resta_delictes ~ offset(log(Pobl16)) + Taxa_envelliment + Estrangers + Turisme + Risc_pobresa + Segona_etapa_E.Secundaria + Divorcis + va_indicadora, data=df, family = negative.binomial(mod7.nb01$theta))
```

Mirem si la variable **Renda_mitjana_llars** és important afegir-la o no:

```
anova(mod6.nb001, mod7.nb001, test = "Chisq")
```

```
## Analysis of Deviance Table
##
## Model 1: Resta_delictes ~ offset(log(Pobl16)) + Taxa_envelliment + Estrangers +
##   Turisme + Renda_mitjana_llars + Risc_pobresa + Segona_etapa_E.Secundaria +
##   Divorcis + va_indicadora
## Model 2: Resta_delictes ~ offset(log(Pobl16)) + Taxa_envelliment + Estrangers +
##   Turisme + Risc_pobresa + Segona_etapa_E.Secundaria + Divorcis +
##   va_indicadora
##   Resid. Df Resid. Dev Df Deviance Pr(>Chi)
## 1     93   102.24
## 2     94   102.23 -1 0.012583
```

```
BIC(mod6.nb001, mod7.nb001)
```

```
##      df    BIC
## mod6.nb001 9 2130.067
## mod7.nb001 8 2126.455
```

La **Renda_mitjana_llars** té influència. Però si ens fixem amb el BIC és menor pel model7.

Anem a veure si podem treure **Aturats**:

```
mod8.nb01 <- glm.nb(Resta_delictes ~ offset(log(Pobl16)) + Taxa_envelliment + Estrangers + Turisme + Risc_pobresa + Segona_etapa_E.Secundaria + Divorcis + va_indicadora, data=df)
```



```
mod8.nb001 <- glm(Resta_delictes ~ offset(log(Pobl16)) + Taxa_envelliment + Estrangers + Turisme + Risc_pobresa + Segona_etapa_E.Secundaria + Divorcis + va_indicadora, data=df, family = negative.binomial(mod8.nb01$theta))
```

Es podria prescindir de la variable **Inferior_2a_etapa_E.Secundaria**?

```
anova(mod7.nb001, mod8.nb001, test = "Chisq")

## Analysis of Deviance Table
##
## Model 1: Resta_delictes ~ offset(log(Pobl16)) + Taxa_envelliment + Estrangers +
##   Turisme + Risc_pobresa + Segona_etapa_E.Secundaria + Divorcis +
##   va_indicadora
## Model 2: Resta_delictes ~ offset(log(Pobl16)) + Taxa_envelliment + Estrangers +
##   Turisme + Risc_pobresa + Segona_etapa_E.Secundaria + Divorcis +
##   va_indicadora
## Resid. Df Resid. Dev Df Deviance Pr(>Chi)
## 1      94    102.23
## 2      94    102.23 0      0

BIC(mod7.nb001, mod8.nb001)

##      df      BIC
## mod7.nb001 8 2126.455
## mod8.nb001 8 2126.455
```

Escollim el model sense la variable, segons el criteri BIC, **Inferior_2a_etapa_E.Secundaria**.

Anem a veure si es podria treure **divorcis**:

```
mod9.nb01 <- glm.nb(Resta_delictes ~ offset(log(Pobl16)) + Taxa_envelliment + Estrangers + Turisme + Risc_pobresa + Segona_etapa_E.Secundaria + va_indicadora, data=df)

mod9.nb001 <- glm(Resta_delictes ~ offset(log(Pobl16)) + Taxa_envelliment + Estrangers + Turisme + Risc_pobresa + Segona_etapa_E.Secundaria + va_indicadora, data=df, family = negative.binomial(mod9.nb01$theta))

anova(mod9.nb001, mod8.nb001, test = "Chisq")

## Analysis of Deviance Table
##
## Model 1: Resta_delictes ~ offset(log(Pobl16)) + Taxa_envelliment + Estrangers +
##   Turisme + Risc_pobresa + Segona_etapa_E.Secundaria + va_indicadora
## Model 2: Resta_delictes ~ offset(log(Pobl16)) + Taxa_envelliment + Estrangers +
##   Turisme + Risc_pobresa + Segona_etapa_E.Secundaria + Divorcis +
##   va_indicadora
## Resid. Df Resid. Dev Df Deviance Pr(>Chi)
## 1      95    102.23
## 2      94    102.23 1 0.0018778 0.9668

BIC(mod7.nb001, mod8.nb001, mod9.nb001)

##      df      BIC
## mod7.nb001 8 2126.455
## mod8.nb001 8 2126.455
## mod9.nb001 7 2124.244
```

```
mod9.nb001 <- stepAIC(mod.nb001, direction = "backward", k = log(nrow(df)), trace = 0) # surt el mateix
que el de adalt perquè s'ha usat el mateix criteri
```

Validació del model

El Análisis del Factor de Inflación de la Varianza (VIF) se utiliza para evaluar la multicolinealidad entre las variables predictoras en un modelo de regresión. Generalmente, valores de VIF inferiores a 5 se consideran aceptables y sugieren que no hay una multicolinealidad grave.

```
# si afegim interacció hauria de ser: vif(mod9.nb001, type = 'predictor')
vif(mod9.nb001)
```

##	Taxa_envelliment	Estrangers	Turisme
##	3.432819	4.068915	1.667720
##	Risc_pobresa	Segona_etapa_E.Secundaria	va_indicadora
##	1.483614	1.546548	1.257102

Todos estos valores están por debajo del umbral de 5, lo que sugiere que no hay una multicolinealidad seria entre las variables predictoras en el modelo. Esto es una buena señal, ya que sugiere que las variables predictoras son relativamente independientes entre sí y no hay problemas graves de colinealidad. En resumen, según el VIF, tus variables parecen estar bien en términos de multicolinealidad en el modelo.

Avaluació del bondat d'ajust del model binomial negatiu:

Contrast d'hipòtesis:

H0: El model és vàlid

H1: El model no és vàlid

```
paste0("L'estadístic de Pearson és ", round(X2P <- sum(resid(mod9.nb001,type="pearson")^2),4)) # estadí
stic de pearson
```

```
## [1] "L'estadístic de Pearson és 111.797"
```

```
paste0("El pvalor és: ", round(1 - pchisq(X2P,mod9.nb001$df.residual),4))
```

```
## [1] "El pvalor és: 0.1149"
```

```
paste0("La deviància del model és: ", round(D <- mod9.nb001$deviance,4))
```

```
## [1] "La deviància del model és: 112.3515"
```

```
paste0("El pvalor del test de la deviància és: ", round(1 - pchisq(D,mod9.nb001$df.residual),4))
```

```
## [1] "El pvalor del test de la deviància és: 0.108"
```

Els p-valors surten superiors al nivell de significació que és 0.05. Usant Dev o X2P no tinc evidència per rebutjar la hipotesi nula: el model ajusta bé les dades.

Proseguim amb la validació gràfica:

```
residualPlot(mod9.nb001)
```

```
residualPlots(mod9.nb001)
```

```
##           Test stat Pr(>|Test stat|)
## Taxa_envelliment      0.5648      0.452328
## Estrangers            0.0829      0.773349
## Turisme              10.0238      0.001545 **
## Risc_pobresa          2.6411      0.104131
## Segona_etapa_E.Secundaria 1.4414      0.229918
## va_indicadora         0.0000      1.000000
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Necessitem posar les variables: **Turisme al quadrat**:

```
mod9.nb01 <- glm.nb(Resta_delictes ~ offset(log(Pobl16)) + Taxa_envelliment + Estrangers + Turisme + I(
Turisme^2) + Risc_pobresa + Segona_etapa_E.Secundaria + va_indicadora, data=df)
```

```
mod9.nb001 <- glm(Resta_delictes ~ offset(log(Pobl16)) + Taxa_envelliment + Estrangers + Turisme + I(Tu
risme^2) + Risc_pobresa + Segona_etapa_E.Secundaria + va_indicadora, data=df, family = negative.binom
ial(mod9.nb01$theta))
```

```
summary(mod9.nb001)
```

```
##
## Call:
## glm(formula = Resta_delictes ~ offset(log(Pobl16)) + Taxa_envelliment +
##   Estrangers + Turisme + I(Turisme^2) + Risc_pobresa + Segona_etapa_E.Secundaria +
##   va_indicadora, family = negative.binomial(mod9.nb01$theta),
##   data = df)
##
## Deviance Residuals:
##   Min     1Q   Median     3Q      Max
## -2.6607 -0.6172 -0.1085  0.4672  2.4195
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    -2.2116079  0.1539126 -14.369 < 2e-16 ***
## Taxa_envelliment    -0.0027645  0.0004979  -5.553 2.60e-07 ***
## Estrangers         2.0071610  0.3879414   5.174 1.29e-06 ***
## Turisme          1.5649173  0.2985028   5.243 9.71e-07 ***
## I(Turisme^2)     -2.0873499  0.6880875  -3.034 0.003125 **
## Risc_pobresa      -0.0175510  0.0017247 -10.176 < 2e-16 ***
## Segona_etapa_E.Secundaria -0.0225917  0.0064485  -3.503 0.000705 ***
## va_indicadora     -0.1038661  0.0240168  -4.325 3.80e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for Negative Binomial(105.3406) family taken to be 1.090181)
##
##   Null deviance: 964.54  on 101  degrees of freedom
## Residual deviance: 102.27  on  94  degrees of freedom
## AIC: 2098.3
##
## Number of Fisher Scoring iterations: 4
```

```
Anova(mod9.nb001) # contribució marginal
```

```

## Analysis of Deviance Table (Type II tests)
##
## Response: Resta_delictes
##          LR Chisq Df Pr(>Chisq)
## Taxa_envelliment      31.104  1 2.445e-08 ***
## Estrangers            27.017  1 2.017e-07 ***
## Turisme               27.486  1 1.582e-07 ***
## l(Turisme^2)           9.190  1 0.0024339 **
## Risc_pobresa          103.313  1 < 2.2e-16 ***
## Segona_etapa_E.Secundaria 12.182  1 0.0004825 ***
## va_indicadora          18.482  1 1.715e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

paste0("La deviància del model és: ", round(D <- mod9.nb001$deviance,4))

## [1] "La deviància del model és: 102.2725"

paste0("El pvalor del test de la deviància és: ", round(1 - pchisq(D,mod9.nb001$df.residual),4))

## [1] "El pvalor del test de la deviància és: 0.2628"

paste0("L'estadístic de Pearson és ", round(X2P <- sum(resid(mod9.nb001,type="pearson")^2),4)) # estadí
stic de pearson

## [1] "L'estadístic de Pearson és 102.477"

paste0("El pvalor és: ", round(1 - pchisq(X2P,mod9.nb001$df.residual),4))

## [1] "El pvalor és: 0.2582"

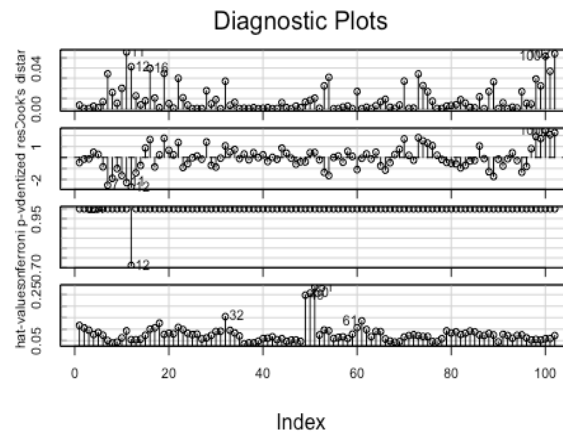
residualPlot(mod9.nb001)

residualPlots(mod9.nb001)

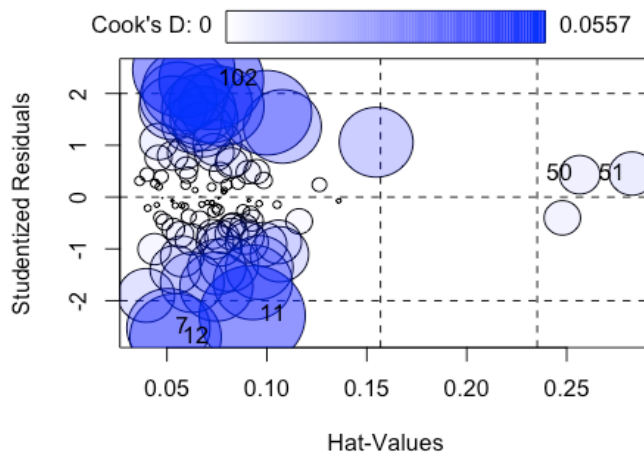
##          Test stat Pr(>|Test stat|)
## Taxa_envelliment      0.6728      0.4121
## Estrangers            0.1493      0.6992
## Turisme               0.0000      1.0000
## l(Turisme^2)           0.3856      0.5346
## Risc_pobresa          2.3421      0.1259
## Segona_etapa_E.Secundaria 0.6562      0.4179
## va_indicadora          0.0000      1.0000

influenceIndexPlot(mod9.nb001,id=list(vars=c("Cook", "hat"), n=5))

```



`influencePlot(mod9.nb001)`

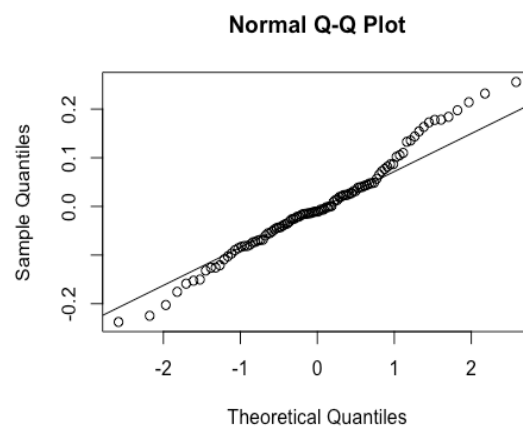


##	StudRes	Hat	CookD
## 7	-2.5243932	0.05072198	0.0342858
56			
## 11	-2.2874760	0.09267472	0.0557006
41			
## 12	-2.6982187	0.05430937	0.0413164
62			
## 50	0.4315820	0.25649277	0.0082383
27			
## 51	0.4535517	0.28264473	0.0103884
73			
## 102	2.2585069	0.07219981	0.0540706
28			

Paquete: base (sin necesidad de paquete adicional)

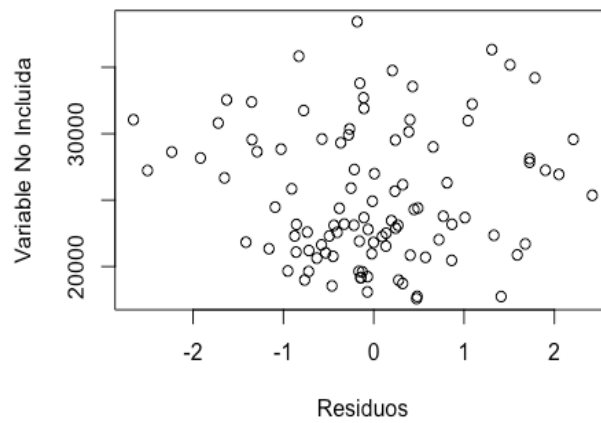
`qqnorm(mod9.nb001$residuals)`

`qqline(mod9.nb001$residuals)` *# Paquete: stats (sin necesidad de paquete adicional)*

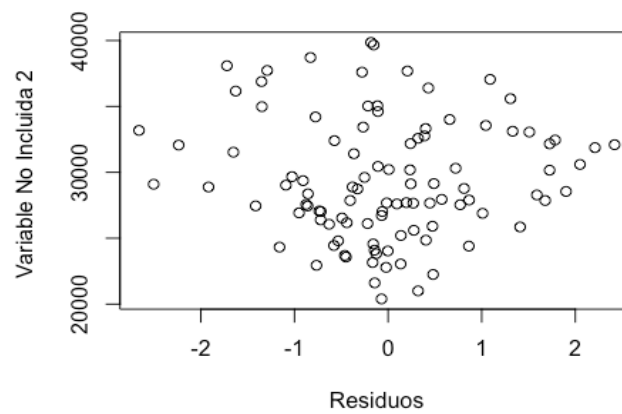


Residus vs variables no incloses al model:

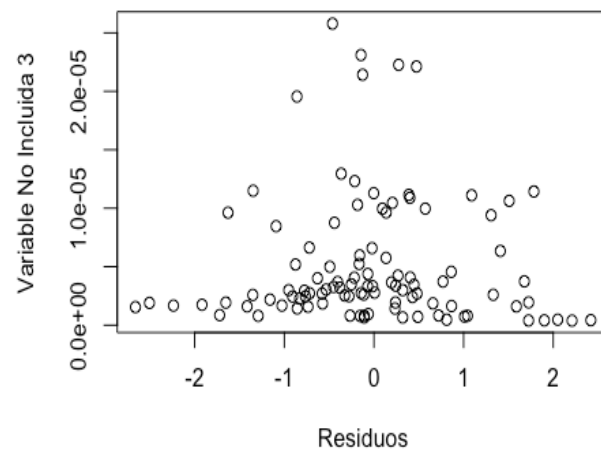
`plot(resid(mod9.nb001), df$PIB_pc, xlab = "Residuos", ylab = "Variable No Incluida")`



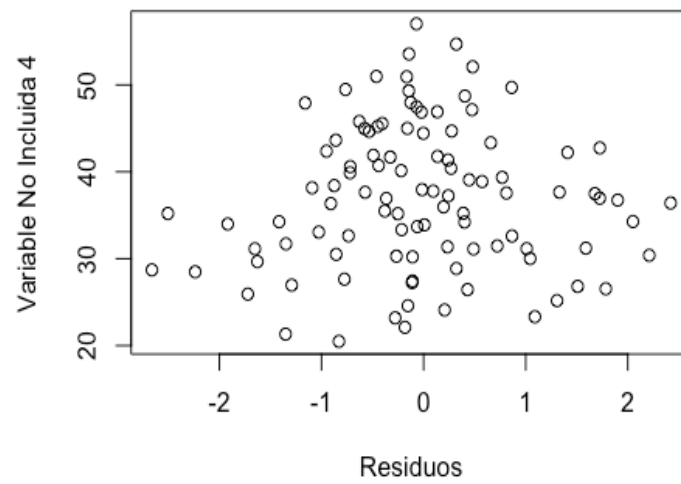
```
plot(resid(mod9.nb001), df$Renda_mitjana_llars, xlab = "Residuos", ylab = "Variable No Incluida 2 ")
```



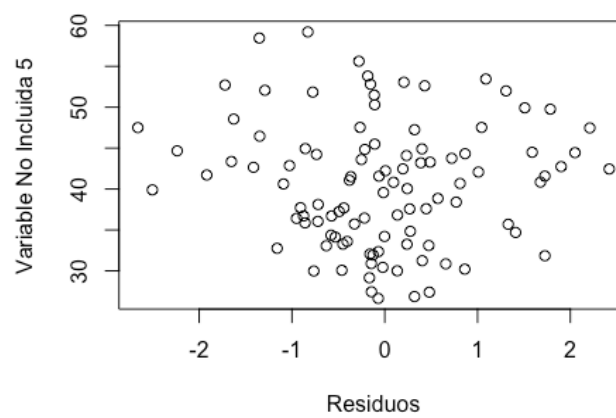
```
plot(resid(mod9.nb001), df$Aturats, xlab = "Residuos", ylab = "Variable No Incluida 3")
```



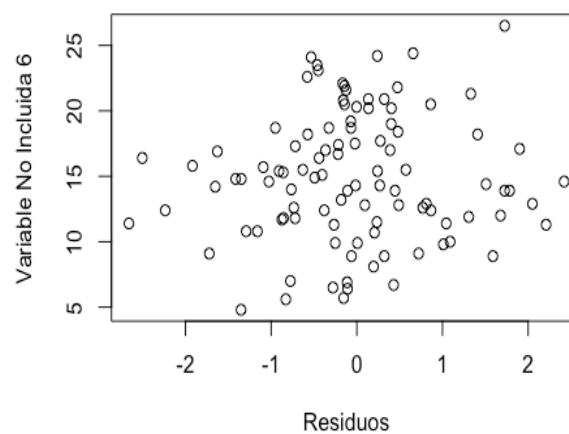
```
plot(resid(mod9.nb001), df$Inferior_2a_etapa_E.Secundaria, xlab = "Residuos", ylab = "Variable No Incluida 4")
```



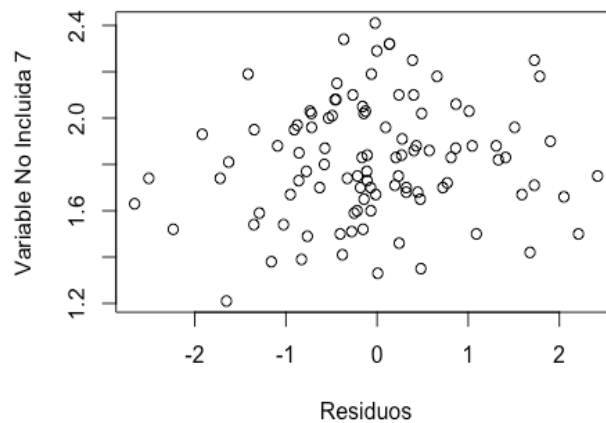
```
plot(resid(mod9.nb001), df$E.Superior, xlab = "Residuos", ylab = "Variable No Incluida 5")
```



```
plot(resid(mod9.nb001), df$Taxa_abandonament_aviat_escolar, xlab = "Residuos", ylab = "Variable No Incluida 6")
```



```
plot(resid(mod9.nb001), df$Divorcis, xlab = "Residuos", ylab = "Variable No Incluida 7")
```



Punts extrems:

```
outlierTest(mod9.nb001)
```

```
## No Studentized residuals with Bonferroni p < 0.05
## Largest |rstudent|:
## rstudent unadjusted p-value Bonferroni p
## 12 -2.698219 0.0069712 0.71106
```

Model 2 - Delictes contra la persona

Model de Poisson

```
dades <- as.data.frame(taxa_factors)
```

```
dades$Anys <- NULL
```

```
dades$CCAA <- NULL
```

```
dades$Resta_delictes <- NULL
```

Realizar la selección de variables hacia atrás con stepAIC

```
modelo_poisson2 <- stepAIC(model2_full, direction = "backward", k = log(nrow(dades)), trace = 0)
```

```
summary(modelo_poisson2)
```

```
## Call:
```

```
## glm(formula = Delictes_contra_persona ~ offset(log(Pobl16)) +
## Taxa_envelliment + Estrangers + Turisme + PIB_pc + Renda_mitjana_llars +
## Risc_pobresa + Aturats + Inferior_2a_etapa_E.Secundaria +
## Segona_etapa_E.Secundaria + E.Superior + Taxa_abandonament_aviat_escolar +
## Divorcis + va_indicadora, family = poisson, data = dades)
##
```

```
## Deviance Residuals:
```

```
## Min 1Q Median 3Q Max
## -36.336 -18.823 -7.176 7.571 64.474
##
```

```
## Coefficients:
```

```
## Estimate Std. Error z value Pr(>|z|)
## (Intercept) 6.482e+04 5.835e+03 11.109 < 2e-16 ***
## Taxa_envelliment -7.936e-03 1.741e-04 -45.581 < 2e-16 ***
## Estrangers 3.857e+00 1.258e-01 30.663 < 2e-16 ***
## Turisme 1.509e-01 3.535e-02 4.269 1.96e-05 ***
## PIB_pc -2.169e-05 2.115e-06 -10.257 < 2e-16 ***
## Renda_mitjana_llars -5.126e-05 2.443e-06 -20.982 < 2e-16 ***
## Risc_pobresa -9.312e-02 9.822e-04 -94.813 < 2e-16 ***
## Aturats -4.073e+04 7.511e+02 -54.223 < 2e-16 ***
## Inferior_2a_etapa_E.Secundaria -6.482e+02 5.835e+01 -11.108 < 2e-16 ***
```



```
## Segona_etapa_E.Secundaria    -6.483e+02  5.835e+01 -11.111 < 2e-16 ***
## E.Superior                   -6.482e+02  5.835e+01 -11.109 < 2e-16 ***
## Taxa_abandonament_aviat_escolar -1.596e-02  1.422e-03 -11.220 < 2e-16 ***
## Divorcis                     -1.276e+00  1.843e-02 -69.214 < 2e-16 ***
## va_indicadora                -3.866e-01  9.244e-03 -41.821 < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for poisson family taken to be 1)
##
## Null deviance: 79945  on 101  degrees of freedom
## Residual deviance: 42721  on  88  degrees of freedom
## AIC: 43606
##
## Number of Fisher Scoring iterations: 5
```

Cal mirar la sobredispersió, que en el cas d'haver-hi, s'estarien infravalorant els errors estàndards, això explicaria perquè surten tants coeficients significatius.

```
# Estimar paràmetre de sobredispersió
X2P <- sum(resid(modelo_poisson2,type="pearson")^2)
phi <- X2P/modelo_poisson2$df.residual
paste0("El paràmetre de sobredispersió estimat és ", round(phi,3))

## [1] "El paràmetre de sobredispersió estimat és 482.725"

# Mirem test de sobredispersió: Trafo = 1 & Trafo = 2:

dispersiontest(modelo_poisson2, trafo = 1) #  $V(Y) = (1+\alpha)*\mu \rightarrow Y$  Quasi Poisson
## Overdispersion test
##
## data: modelo_poisson2
## z = 5.2766, p-value = 6.58e-08
## alternative hypothesis: true alpha is greater than 0
## sample estimates:
## alpha
## 415.6024

dispersiontest(modelo_poisson2, trafo = 2) #  $V(Y) = \mu + \alpha*\mu^2 \rightarrow Y$  Bin Neg
## Overdispersion test
##
## data: modelo_poisson2
## z = 4.4062, p-value = 5.26e-06
## alternative hypothesis: true alpha is greater than 0
## sample estimates:
## alpha
## 0.1883044
```

Com trafo = 1 i trafo =2 són significatius, la relació entre l'esperança i la variància és quadràtica.

Model Binomial Negatiu

```
mod2.nb01 <- glm.nb(Delictes_contra_persona ~ offset(log(Pobl16)) + Taxa_envelliment + Estrangers +
PIB_pc + Renda_mitjana_llars + Risc_pobresa + Aturats + Inferior_2a_etapa_E.Secundaria + Segona_etapa
_E.Secundaria + E.Superior + Taxa_abandonament_aviat_escolar + Divorcis + va_indicadora, data=dades)

mod2.nb001 <- glm(Delictes_contra_persona ~ offset(log(Pobl16)) + Taxa_envelliment + Estrangers + PIB
```

```
_pc + Renda_mitjana_llars + Risc_pobresa + Aturats + Inferior_2a_etapa_E.Secundaria + Segona_etapa_E.
Secundaria + E.Superior + Taxa_abandonament_aviat_escolar + Divorcis + va_indicadora, data=dades, fam
ily = negative.binomial(mod2.nb01$theta))
```

```
summary(mod2.nb001)
```

```
##
## Call:
## glm(formula = Delictes_contra_persona ~ offset(log(Pobl16)) +
## Taxa_envelliment + Estrangers + PIB_pc + Renda_mitjana_llars +
## Risc_pobresa + Aturats + Inferior_2a_etapa_E.Secundaria +
## Segona_etapa_E.Secundaria + E.Superior + Taxa_abandonament_aviat_escolar +
## Divorcis + va_indicadora, family = negative.binomial(mod2.nb01$theta),
## data = dades)
##
## Deviance Residuals:
## Min 1Q Median 3Q Max
## -2.5277 -0.8398 -0.3138 0.4069 2.5393
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.061e+04 1.710e+05 0.062 0.950678
## Taxa_envelliment -1.067e-02 4.072e-03 -2.621 0.010306 *
## Estrangers 7.890e+00 3.157e+00 2.499 0.014294 *
## PIB_pc 1.546e-05 5.167e-05 0.299 0.765424
## Renda_mitjana_llars -7.152e-05 6.249e-05 -1.144 0.255535
## Risc_pobresa -9.409e-02 2.489e-02 -3.781 0.000282 ***
## Aturats -6.152e+03 1.872e+04 -0.329 0.743143
## Inferior_2a_etapa_E.Secundaria -1.060e+02 1.710e+03 -0.062 0.950717
## Segona_etapa_E.Secundaria -1.062e+02 1.710e+03 -0.062 0.950632
## E.Superior -1.061e+02 1.710e+03 -0.062 0.950680
## Taxa_abandonament_aviat_escolar -5.997e-02 3.863e-02 -1.552 0.124130
## Divorcis -1.843e+00 4.326e-01 -4.259 5.08e-05 ***
## va_indicadora -3.979e-01 2.512e-01 -1.584 0.116825
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for Negative Binomial(2.0303) family taken to be 1.251305)
##
## Null deviance: 170.07 on 101 degrees of freedom
## Residual deviance: 110.25 on 89 degrees of freedom
## AIC: 1600.9
##
## Number of Fisher Scoring iterations: 25
```

```
mod2.nb02 <- glm.nb(Delictes_contra_persona ~ offset(log(Pobl16)) + Taxa_envelliment + Estrangers +
PIB_pc + Renda_mitjana_llars + Risc_pobresa + Aturats + Segona_etapa_E.Secundaria + E.Superior + Taxa_
abandonament_aviat_escolar + Divorcis + va_indicadora, data=dades)
```

```
mod2.nb002 <- glm(Delictes_contra_persona ~ offset(log(Pobl16)) + Taxa_envelliment + Estrangers + PIB
_pc + Renda_mitjana_llars + Risc_pobresa + Aturats + Segona_etapa_E.Secundaria + E.Superior + Taxa_ab
andonament_aviat_escolar + Divorcis + va_indicadora, data=dades, family = negative.binomial(mod2.nb0
2$theta))
```

```
anova(mod2.nb001, mod2.nb002, test='Chisq')
```

```
## Analysis of Deviance Table
##
## Model 1: Delictes_contra_persona ~ offset(log(Pobl16)) + Taxa_envelliment +
##   Estrangers + PIB_pc + Renda_mitjana_llars + Risc_pobresa +
##   Aturats + Inferior_2a_etapa_E.Secundaria + Segona_etapa_E.Secundaria +
##   E.Superior + Taxa_abandonament_aviat_escolar + Divorcis +
##   va_indicadora
## Model 2: Delictes_contra_persona ~ offset(log(Pobl16)) + Taxa_envelliment +
##   Estrangers + PIB_pc + Renda_mitjana_llars + Risc_pobresa +
##   Aturats + Segona_etapa_E.Secundaria + E.Superior + Taxa_abandonament_aviat_escolar +
##   Divorcis + va_indicadora
## Resid. Df Resid. Dev Df   Deviance Pr(>Chi)
## 1      89   110.25
## 2      90   110.25 -1 -0.0001238  0.9921
```

Treiem la variable: **Inferior_2a_etapa_E.Secundaria**

```
mod2.nb03 <- glm.nb(Delictes_contra_persona ~ offset(log(Pobl16)) + Taxa_envelliment + Estrangers +
Renda_mitjana_llars + Risc_pobresa + Aturats + Segona_etapa_E.Secundaria + E.Superior + Taxa_abandonament_aviat_escolar + Divorcis + va_indicadora, data=dades)
```

```
mod2.nb003 <- glm(Delictes_contra_persona ~ offset(log(Pobl16)) + Taxa_envelliment + Estrangers + Renda_mitjana_llars + Risc_pobresa + Aturats + Segona_etapa_E.Secundaria + E.Superior + Taxa_abandonament_aviat_escolar + Divorcis + va_indicadora, data=dades, family = negative.binomial(mod2.nb03$theta))
```

```
anova(mod2.nb002, mod2.nb003, test='Chisq')
```

```
## Analysis of Deviance Table
##
## Model 1: Delictes_contra_persona ~ offset(log(Pobl16)) + Taxa_envelliment +
##   Estrangers + PIB_pc + Renda_mitjana_llars + Risc_pobresa +
##   Aturats + Segona_etapa_E.Secundaria + E.Superior + Taxa_abandonament_aviat_escolar +
##   Divorcis + va_indicadora
## Model 2: Delictes_contra_persona ~ offset(log(Pobl16)) + Taxa_envelliment +
##   Estrangers + Renda_mitjana_llars + Risc_pobresa + Aturats +
##   Segona_etapa_E.Secundaria + E.Superior + Taxa_abandonament_aviat_escolar +
##   Divorcis + va_indicadora
## Resid. Df Resid. Dev Df   Deviance Pr(>Chi)
## 1      90   110.25
## 2      91   110.27 -1 -0.013643  0.9164
```

Podem treure la variable **Aturats**. Seguim reduint el nombre de variables.

```
mod2.nb04 <- glm.nb(Delictes_contra_persona ~ offset(log(Pobl16)) + Taxa_envelliment + Estrangers +
Renda_mitjana_llars + Risc_pobresa + Segona_etapa_E.Secundaria + E.Superior + Taxa_abandonament_aviat_escolar + Divorcis + va_indicadora, data=dades)
```

```
mod2.nb004 <- glm(Delictes_contra_persona ~ offset(log(Pobl16)) + Taxa_envelliment + Estrangers + Renda_mitjana_llars + Risc_pobresa + Segona_etapa_E.Secundaria + E.Superior + Taxa_abandonament_aviat_escolar + Divorcis + va_indicadora, data=dades, family = negative.binomial(mod2.nb04$theta))
```

```
anova(mod2.nb003, mod2.nb004, test='Chisq')
```

```
## Analysis of Deviance Table
##
## Model 1: Delictes_contra_persona ~ offset(log(Pobl16)) + Taxa_envelliment +
```

```
## Estrangers + Renda_mitjana_llars + Risc_pobresa + Aturats +
## Segona_etapa_E.Secundaria + E.Superior + Taxa_abandonament_aviat_escolar +
## Divorcis + va_indicadora
## Model 2: Delictes_contra_persona ~ offset(log(Pobl16)) + Taxa_envelliment +
## Estrangers + Renda_mitjana_llars + Risc_pobresa + Segona_etapa_E.Secundaria +
## E.Superior + Taxa_abandonament_aviat_escolar + Divorcis +
## va_indicadora
## Resid. Df Resid. Dev Df Deviance Pr(>Chi)
## 1 91 110.27
## 2 92 110.26 -1 0.0077848
```

Hi ha evidència estadística per a preferir el model 2 sobre el model 1.

Treiem variable renda mitjana de llars:

```
mod2.nb05 <- glm.nb(Delictes_contra_persona ~ offset(log(Pobl16)) + Taxa_envelliment + Estrangers +
Risc_pobresa + Segona_etapa_E.Secundaria + E.Superior + Taxa_abandonament_aviat_escolar + Divorcis +
va_indicadora, data=dades)
```

```
mod2.nb005 <- glm(Delictes_contra_persona ~ offset(log(Pobl16)) + Taxa_envelliment + Estrangers + Risc_pobresa + Segona_etapa_E.Secundaria + E.Superior + Taxa_abandonament_aviat_escolar + Divorcis + va_indicadora, data=dades, family = negative.binomial(mod2.nb05$theta))
```

```
anova(mod2.nb004, mod2.nb005, test='Chisq')
```

```
## Analysis of Deviance Table
```

```
##
```

```
## Model 1: Delictes_contra_persona ~ offset(log(Pobl16)) + Taxa_envelliment +
## Estrangers + Renda_mitjana_llars + Risc_pobresa + Segona_etapa_E.Secundaria +
## E.Superior + Taxa_abandonament_aviat_escolar + Divorcis +
## va_indicadora
```

```
## Model 2: Delictes_contra_persona ~ offset(log(Pobl16)) + Taxa_envelliment +
## Estrangers + Risc_pobresa + Segona_etapa_E.Secundaria + E.Superior +
## Taxa_abandonament_aviat_escolar + Divorcis + va_indicadora
```

```
## Resid. Df Resid. Dev Df Deviance Pr(>Chi)
```

```
## 1 92 110.26
```

```
## 2 93 110.37 -1 -0.10759 0.7664
```

Treiem taxa abandonament aviat escolar:

```
mod2.nb06 <- glm.nb(Delictes_contra_persona ~ offset(log(Pobl16)) + Taxa_envelliment + Estrangers +
Risc_pobresa + Segona_etapa_E.Secundaria + E.Superior + Divorcis + va_indicadora, data=dades)
```

```
mod2.nb006 <- glm(Delictes_contra_persona ~ offset(log(Pobl16)) + Taxa_envelliment + Estrangers + Risc_pobresa + Segona_etapa_E.Secundaria + E.Superior + Divorcis + va_indicadora, data=dades, family = negative.binomial(mod2.nb06$theta))
```

```
anova(mod2.nb005, mod2.nb006, test='Chisq')
```

```
## Analysis of Deviance Table
```

```
##
```

```
## Model 1: Delictes_contra_persona ~ offset(log(Pobl16)) + Taxa_envelliment +
## Estrangers + Risc_pobresa + Segona_etapa_E.Secundaria + E.Superior +
## Taxa_abandonament_aviat_escolar + Divorcis + va_indicadora
```

```
## Model 2: Delictes_contra_persona ~ offset(log(Pobl16)) + Taxa_envelliment +
## Estrangers + Risc_pobresa + Segona_etapa_E.Secundaria + E.Superior +
## Divorcis + va_indicadora
```

```
## Resid. Df Resid. Dev Df Deviance Pr(>Chi)
## 1    93   110.37
## 2    94   110.61 -1 -0.24557  0.6621

# Treiem estrangers:
mod2.nb07 <- glm.nb(Delictes_contra_persona ~ offset(log(Pobl16)) + Taxa_envelliment + Risc_pobresa + Segona_etapa_E.Secundaria + E.Superior + Divorcis + va_indicadora, data=dades)

mod2.nb007 <- glm(Delictes_contra_persona ~ offset(log(Pobl16)) + Taxa_envelliment + Risc_pobresa + Segona_etapa_E.Secundaria + E.Superior + Divorcis + va_indicadora, data=dades, family = negative.binomial(mod2.nb07$theta))

anova(mod2.nb006, mod2.nb007, test='Chisq')

## Analysis of Deviance Table
##
## Model 1: Delictes_contra_persona ~ offset(log(Pobl16)) + Taxa_envelliment +
##   Estrangers + Risc_pobresa + Segona_etapa_E.Secundaria + E.Superior +
##   Divorcis + va_indicadora
## Model 2: Delictes_contra_persona ~ offset(log(Pobl16)) + Taxa_envelliment +
##   Risc_pobresa + Segona_etapa_E.Secundaria + E.Superior + Divorcis +
##   va_indicadora
## Resid. Df Resid. Dev Df Deviance Pr(>Chi)
## 1    94   110.61
## 2    95   110.77 -1 -0.15616  0.723

# Treiem va_indicadora:
mod2.nb08 <- glm.nb(Delictes_contra_persona ~ offset(log(Pobl16)) + Taxa_envelliment + Risc_pobresa + Segona_etapa_E.Secundaria + E.Superior + Divorcis, data=dades)

mod2.nb008 <- glm(Delictes_contra_persona ~ offset(log(Pobl16)) + Taxa_envelliment + Risc_pobresa + Segona_etapa_E.Secundaria + E.Superior + Divorcis, data=dades, family = negative.binomial(mod2.nb08$theta))

summary(mod2.nb008)

##
## Call:
## glm(formula = Delictes_contra_persona ~ offset(log(Pobl16)) +
##   Taxa_envelliment + Risc_pobresa + Segona_etapa_E.Secundaria +
##   E.Superior + Divorcis, family = negative.binomial(mod2.nb08$theta),
##   data = dades)
##
## Deviance Residuals:
##   Min     1Q   Median     3Q      Max
## -2.3247 -0.8744 -0.3185  0.3752  2.7633
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    4.457689   1.850411   2.409 0.017903 *
## Taxa_envelliment -0.013521  0.002426 -5.572 2.30e-07 ***
## Risc_pobresa    -0.082554  0.019924 -4.143 7.37e-05 ***
## Segona_etapa_E.Secundaria -0.132202  0.047466 -2.785 0.006445 **
## E.Superior     -0.067836  0.018506 -3.666 0.000405 ***
## Divorcis       -1.448728  0.363063 -3.990 0.000129 ***
```

```
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for Negative Binomial(1.8559) family taken to be 1.260201)
##
## Null deviance: 155.59 on 101 degrees of freedom
## Residual deviance: 110.97 on 96 degrees of freedom
## AIC: 1597.4
##
## Number of Fisher Scoring iterations: 14

anova(mod2.nb007, mod2.nb008, test='Chisq')

## Analysis of Deviance Table
##
## Model 1: Delictes_contra_persona ~ offset(log(Pobl16)) + Taxa_envelliment +
## Risc_pobresa + Segona_etapa_E.Secundaria + E.Superior + Divorcis +
## va_indicadora
## Model 2: Delictes_contra_persona ~ offset(log(Pobl16)) + Taxa_envelliment +
## Risc_pobresa + Segona_etapa_E.Secundaria + E.Superior + Divorcis
## Resid. Df Resid. Dev Df Deviance Pr(>Chi)
## 1 95 110.77
## 2 96 110.97 -1 -0.20325 0.6851
```

Si usem la funció STEP arribem al mateix model (mod2.nb008):

```
adjusted_model <- stepAIC(mod2.nb001, direction = "backward", k = log(nrow(dades)), trace = 0)

summary(adjusted_model)
```

Validació del model

```
vif(mod2.nb008)
```

##	Taxa_envelliment	Risc_pobresa	Segona_etapa_E.Secundaria
##	1.237435	3.281052	1.294185
##	E.Superior	Divorcis	
##	2.968247	1.244181	

No tenim problemes de multicolinealitat.

```
# Pearson:
paste0("L'estadístic de Pearson és ", round(X2P <- sum(resid(mod2.nb008,type="pearson")^2),4))

## [1] "L'estadístic de Pearson és 120.9793"

paste0("El pvalor és: ", round(1 - pchisq(X2P,mod2.nb008$df.residual),4))

## [1] "El pvalor és: 0.0433"
```

Per Pearson no quedaria validat el model.

```
# Deviància:
paste0("La deviància del model és: ", round(D <- mod2.nb008$deviance,4))

## [1] "La deviància del model és: 110.9716"

paste0("El pvalor del test de la deviància és: ", round(1 - pchisq(D,mod2.nb008$df.residual),4))
```

```
## [1] "El pvalor del test de la deviancia és: 0.1409"
```

Pel test de la Deviancia quedaria validat el model.

```
residualPlot(mod2.nb008)
```

```
residualPlots(mod2.nb008)
```

```
##           Test stat Pr(>|Test stat|)
## Taxa_envelliment      0.2487      0.617969
## Risc_pobresa          2.0937      0.147904
## Segona_etapa_E.Secundaria  7.5839      0.005889 **
## E.Superior             0.6865      0.407359
## Divorcis               0.0361      0.849316
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
outlierTest(mod2.nb008)
```

```
## No Studentized residuals with Bonferroni p < 0.05
## Largest |rstudent|:
##  rstudent unadjusted p-value Bonferroni p
## 39 2.942119      0.0032597  0.33249
```

Afegim terme quadràtic a Segona_etapa_E.Secundaria:

```
mod2.nb09 <- glm.nb(Delictes_contra_persona ~ offset(log(Pobl16)) + Taxa_envelliment + Risc_pobresa
+ Segona_etapa_E.Secundaria + I(Segona_etapa_E.Secundaria^2) + E.Superior + Divorcis, data=dades)
```

```
mod2.nb009 <- glm(Delictes_contra_persona ~ offset(log(Pobl16)) + Taxa_envelliment + Risc_pobresa + S
egona_etapa_E.Secundaria + I(Segona_etapa_E.Secundaria^2) + E.Superior + Divorcis, data=dades, family
= negative.binomial(mod2.nb09$theta))
```

```
summary(mod2.nb009)
```

```
##
## Call:
## glm(formula = Delictes_contra_persona ~ offset(log(Pobl16)) +
##  Taxa_envelliment + Risc_pobresa + Segona_etapa_E.Secundaria +
##  I(Segona_etapa_E.Secundaria^2) + E.Superior + Divorcis, family = negative.binomial(mod2.nb09$the
ta),
##  data = dades)
##
## Deviance Residuals:
##   Min     1Q   Median     3Q      Max
## -2.3889 -0.9168 -0.3278  0.4152  2.6461
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    -17.48889   8.09063  -2.162 0.033160 *
## Taxa_envelliment    -0.01462   0.00234  -6.248 1.17e-08 ***
## Risc_pobresa       -0.07861   0.01913  -4.108 8.46e-05 ***
## Segona_etapa_E.Secundaria   1.84450   0.72810   2.533 0.012937 *
## I(Segona_etapa_E.Secundaria^2) -0.04335   0.01602  -2.706 0.008071 **
## E.Superior        -0.06927   0.01809  -3.828 0.000231 ***
## Divorcis         -1.61113   0.35459  -4.544 1.63e-05 ***
```

```
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for Negative Binomial(1.985) family taken to be 1.236891)
##
## Null deviance: 166.31 on 101 degrees of freedom
## Residual deviance: 110.50 on 95 degrees of freedom
## AIC: 1591.6
##
## Number of Fisher Scoring iterations: 12

residualPlot(mod2.nb009)

residualPlots(mod2.nb009)

##              Test stat Pr(>|Test stat|)
## Taxa_envelliment      0.1624      0.68695
## Risc_pobresa          0.3111      0.57700
## Segona_etapa_E.Secundaria 0.0000      1.00000
## I(Segona_etapa_E.Secundaria^2) 1.0481      0.30594
## E.Superior            4.0825      0.04333 *
## Divorcis              0.0021      0.96387
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

# Afegim terme quadrat a E.Superior:

mod2.nb10 <- glm.nb(Delictes_contra_persona ~ offset(log(Pobl16)) + Taxa_envelliment + Risc_pobresa
+ Segona_etapa_E.Secundaria + I(Segona_etapa_E.Secundaria^2) + E.Superior + I(E.Superior^2) + Divorcis,
data=dades)

mod2.nb010 <- glm(Delictes_contra_persona ~ offset(log(Pobl16)) + Taxa_envelliment + Risc_pobresa + S
egona_etapa_E.Secundaria + I(Segona_etapa_E.Secundaria^2) + E.Superior + I(E.Superior^2) + Divorcis, da
ta=dades, family = negative.binomial(mod2.nb10$theta))

summary(mod2.nb010)
## Call:
## glm(formula = Delictes_contra_persona ~ offset(log(Pobl16)) +
## Taxa_envelliment + Risc_pobresa + Segona_etapa_E.Secundaria +
## I(Segona_etapa_E.Secundaria^2) + E.Superior + I(E.Superior^2) +
## Divorcis, family = negative.binomial(mod2.nb10$theta), data = dades)
##
## Deviance Residuals:
##   Min       1Q   Median       3Q      Max
## -2.4651 -0.8654 -0.3553  0.4965  2.7160
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    -19.408462   7.878677  -2.463  0.01558 *
## Taxa_envelliment    -0.013561   0.002359  -5.749 1.11e-07 ***
## Risc_pobresa       -0.084473   0.018878  -4.475 2.14e-05 ***
## Segona_etapa_E.Secundaria    2.415819   0.762708   3.167  0.00207 **
## I(Segona_etapa_E.Secundaria^2) -0.055007   0.016645  -3.305  0.00135 **
## E.Superior        -0.315978   0.131597  -2.401  0.01831 *
## I(E.Superior^2)     0.002786   0.001497   1.861  0.06581 .
```



```
## Divorcis          -1.478127  0.346546 -4.265 4.75e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for Negative Binomial(2.0575) family taken to be 1.201006)
##
## Null deviance: 172.32 on 101 degrees of freedom
## Residual deviance: 110.26 on 94 degrees of freedom
## AIC: 1589.4
##
## Number of Fisher Scoring iterations: 14
```

```
residualPlot(mod2.nb010)
```

```
residualPlots(mod2.nb010)
```

```
##              Test stat Pr(>|Test stat|)
## Taxa_envelliment      0.0320      0.8581
## Risc_pobresa          1.3787      0.2403
## Segona_etapa_E.Secundaria 0.0000      1.0000
## I(Segona_etapa_E.Secundaria^2) 1.2309      0.2672
## E.Superior            0.0000      1.0000
## I(E.Superior^2)        18.2713      1.916e-05 ***
## Divorcis              0.2499      0.6171
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Com surt molt significativa el terme quadràtic d'E.Superior, cal afegir un terme cúbic.

```
mod2.nb11 <- glm.nb(Delictes_contra_persona ~ offset(log(Pobl16)) + Taxa_envelliment + Risc_pobresa +
Segona_etapa_E.Secundaria + I(Segona_etapa_E.Secundaria^2) + E.Superior + I(E.Superior^2) + I(E.Superior
^3) + Divorcis, data=dades)
```

```
mod2.nb011 <- glm(Delictes_contra_persona ~ offset(log(Pobl16)) + Taxa_envelliment + Risc_pobresa + S
egona_etapa_E.Secundaria + I(Segona_etapa_E.Secundaria^2) + E.Superior + I(E.Superior^2) + I(E.Superior
^3) + Divorcis, data=dades, family = negative.binomial(mod2.nb11$theta))
```

```
summary(mod2.nb011)
```

```
## Call:
## glm(formula = Delictes_contra_persona ~ offset(log(Pobl16)) +
## Taxa_envelliment + Risc_pobresa + Segona_etapa_E.Secundaria +
## I(Segona_etapa_E.Secundaria^2) + E.Superior + I(E.Superior^2) +
## I(E.Superior^3) + Divorcis, family = negative.binomial(mod2.nb11$theta),
## data = dades)
##
## Deviance Residuals:
##   Min       1Q   Median       3Q      Max
## -2.6923 -0.9050 -0.2873  0.5033  2.6459
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    -4.755e+01  9.744e+00 -4.880 4.37e-06 ***
## Taxa_envelliment    -1.341e-02  2.135e-03 -6.278 1.08e-08 ***
## Risc_pobresa        -9.195e-02  1.703e-02 -5.399 5.12e-07 ***
## Segona_etapa_E.Secundaria  1.155e+00  7.145e-01  1.616 0.109409
```

```
## l(Segona_etapa_E.Secundaria^2) -2.832e-02 1.557e-02 -1.819 0.072149 .
## E.Superior          2.793e+00 6.646e-01 4.202 6.07e-05 ***
## l(E.Superior^2)      -7.094e-02 1.578e-02 -4.494 2.01e-05 ***
## l(E.Superior^3)      5.688e-04 1.230e-04 4.623 1.22e-05 ***
## Divorcis            -1.254e+00 3.131e-01 -4.004 0.000125 ***
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for Negative Binomial(2.4574) family taken to be 1.16436)
##
## Null deviance: 205.43 on 101 degrees of freedom
## Residual deviance: 109.07 on 93 degrees of freedom
## AIC: 1571
## Number of Fisher Scoring iterations: 12
```

```
residualPlot(mod2.nb011)
```

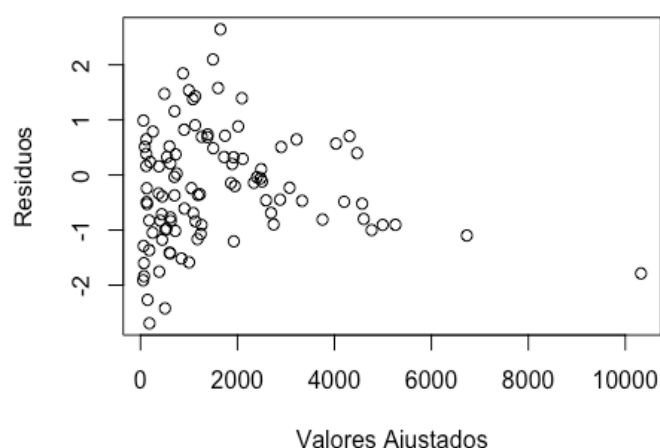
```
residualPlots(mod2.nb011)
```

```
##              Test stat Pr(>|Test stat|)
## Taxa_envelliment      0.5310      0.4662
## Risc_pobresa          0.1895      0.6633
## Segona_etapa_E.Secundaria 0.0000      1.0000
## l(Segona_etapa_E.Secundaria^2) 0.0603      0.8060
## E.Superior            0.0000      1.0000
## l(E.Superior^2)        1.3293      0.2489
## l(E.Superior^3)        1.6593      0.1977
## Divorcis              0.7196      0.3963
```

```
BIC(mod2.nb011, mod2.nb010)
```

```
##      df      BIC
## mod2.nb011 9 1594.632
## mod2.nb010 8 1610.409
```

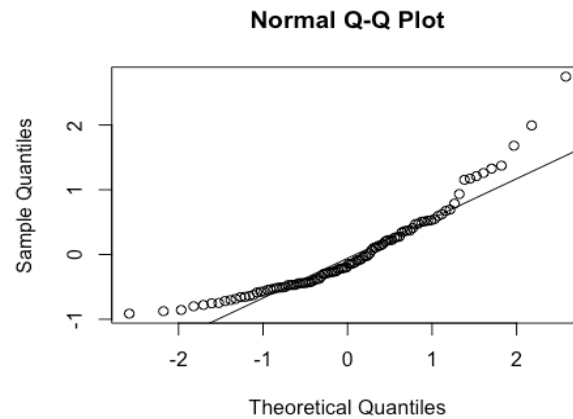
```
plot(mod2.nb011$fit, resid(mod2.nb011), xlab = "Valores Ajustados", ylab = "Residuos")
```



```
# Paquete: base (sin necesidad de paquete adicional)
```

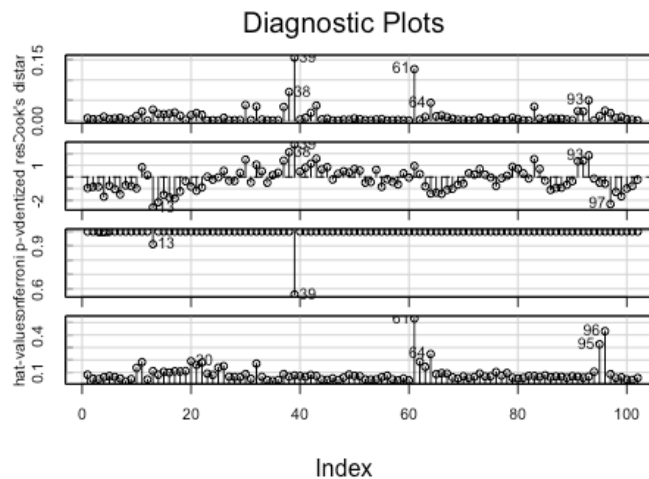
```
qqnorm(mod2.nb011$residuals)
```

```
qqline(mod2.nb011$residuals)
```

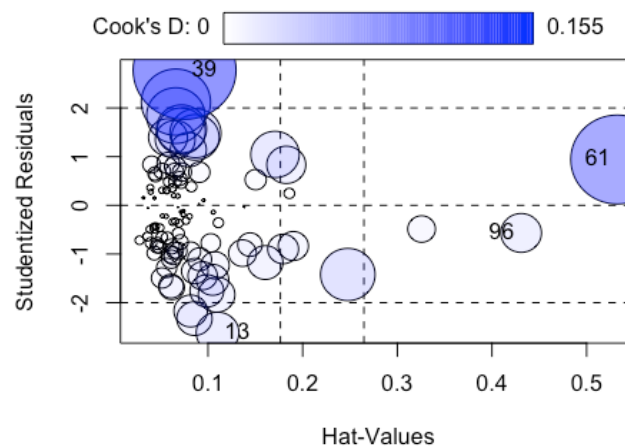


Del gráfico de residuos vs valores ajustados: indica una posible heterocedasticidad o no constancia en la varianza de los errores. Se podría recurrir a transformar las variables.

```
influenceIndexPlot(mod2.nb011,id=list(vars=c("Cook", "hat"), n=5))
```



```
influencePlot(mod2.nb011)
```



```
## StudRes Hat CookD
## 13 -2.6144614 0.11009941 0.02685675
## 39 2.7768065 0.07525346 0.15546550
## 61 0.9409331 0.53236898 0.12695206
## 96 -0.5672250 0.43083888 0.02420201
```

```

# Vector de índices de las filas a eliminar: outliers
indices_a_eliminar <- c(13, 39, 61, 96)

# Elimina las filas con los índices especificados
dades_sense_files <- dades[-indices_a_eliminar, ]
mod2.nb11 <- glm.nb(Delictes_contra_persona ~ offset(log(Pobl16)) + Taxa_envelliment + Risc_pobresa +
Segona_etapa_E.Secundaria + I(Segona_etapa_E.Secundaria^2) + E.Superior + I(E.Superior^2) + I(E.Superior
^3) + Divorcis, data= dades_sense_files)

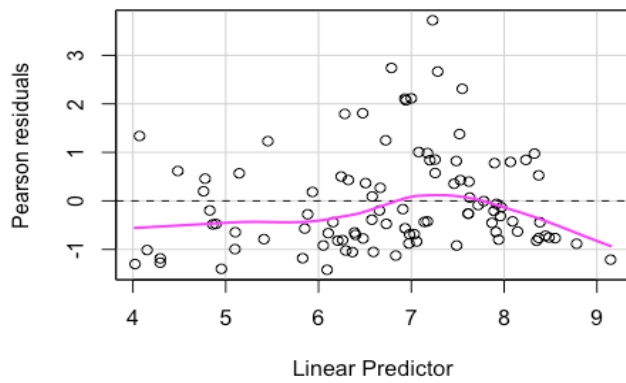
mod2.nb011 <- glm(Delictes_contra_persona ~ offset(log(Pobl16)) + Taxa_envelliment + Risc_pobresa + S
egona_etapa_E.Secundaria + I(Segona_etapa_E.Secundaria^2) + E.Superior + I(E.Superior^2) + I(E.Superior
^3) + Divorcis, data= dades_sense_files, family = negative.binomial(mod2.nb11$theta))

summary(mod2.nb011)

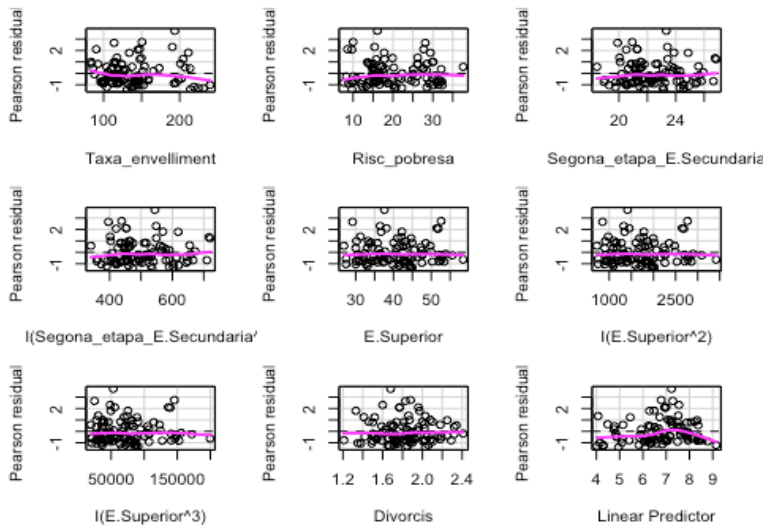
##
## Call:
## glm(formula = Delictes_contra_persona ~ offset(log(Pobl16)) +
## Taxa_envelliment + Risc_pobresa + Segona_etapa_E.Secundaria +
## I(Segona_etapa_E.Secundaria^2) + E.Superior + I(E.Superior^2) +
## I(E.Superior^3) + Divorcis, family = negative.binomial(mod2.nb11$theta),
## data = dades_sense_files)
##
## Deviance Residuals:
## Min 1Q Median 3Q Max
## -2.4360 -0.9333 -0.3143 0.5056 2.4244
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) -5.337e+01 1.231e+01 -4.335 3.82e-05 ***
## Taxa_envelliment -1.362e-02 2.163e-03 -6.296 1.13e-08 ***
## Risc_pobresa -8.287e-02 1.663e-02 -4.984 3.04e-06 ***
## Segona_etapa_E.Secundaria 1.484e+00 8.896e-01 1.669 0.098695 .
## I(Segona_etapa_E.Secundaria^2) -3.513e-02 1.930e-02 -1.820 0.072134 .
## E.Superior 2.910e+00 7.188e-01 4.049 0.000110 ***
## I(E.Superior^2) -7.413e-02 1.739e-02 -4.264 4.99e-05 ***
## I(E.Superior^3) 5.984e-04 1.380e-04 4.337 3.80e-05 ***
## Divorcis -1.153e+00 3.078e-01 -3.746 0.000318 ***
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for Negative Binomial(2.7581) family taken to be 1.177463)
##
## Null deviance: 199.60 on 97 degrees of freedom
## Residual deviance: 104.14 on 89 degrees of freedom
## AIC: 1500.6
##
## Number of Fisher Scoring iterations: 11

residualPlot(mod2.nb011)

```



```
residualPlots(mod2.nb011)
```



##	Test stat	Pr(> Test stat)
## Taxa_envelliment	0.5861	0.4439
## Risc_pobresa	0.0491	0.8247
## Segona_etapa_E.Secundaria	0.0000	1.0000
## l(Segona_etapa_E.Secundaria^2)	1.5945	0.2067
## E.Superior	0.0000	1.0000
## l(E.Superior^2)	1.3987	0.2369
## l(E.Superior^3)	1.6556	0.1982
## Divorcis	0.2302	0.6314