



Universidade de Vigo

Trabajo Fin de Máster

Aplicación de herramientas estadísticas para el análisis de la plasticidad cerebral adulta

Adrián Varela Rodríguez

Máster en Técnicas Estadísticas
Curso 2025-2026

Propuesta de Trabajo Fin de Máster

<p>Título en galego: Aplicación de ferramentas estatísticas para a análise da plasticidade cerebral adulta</p>
<p>Título en español: Aplicación de herramientas estadísticas para el análisis de la plasticidad cerebral adulta</p>
<p>English title: Application of statistical tools for the analysis of adult brain plasticity</p>
<p>Modalidad: Modalidad B</p>
<p>Autor/a: Adrián Varela Rodríguez, Universidade de Santiago de Compostela</p>
<p>Tutor/a académico/a: María Isabel Borrajo García, Universidade de Santiago de Compostela</p>
<p>Tutor/a de empresa: Alba Vieites Prado, CiMUS</p>
<p>Breve resumen del trabajo:</p> <p>Este Trabajo Fin de Máster tiene como objetivo principal analizar la plasticidad cerebral adulta a través de un modelo experimental en ratones, evaluando si el cerebro es capaz de reorganizarse y adaptarse tras la eliminación de las vibrisas faciales. Para ello, se realizará una inyección viral en la principal región de procesamiento táctil de las vibrisas, evaluando la comparabilidad de las inyecciones entre los distintos animales (etapa de validación), tanto de la distribución del número de células en cada capa como teniendo en cuenta la naturaleza espacial de la misma. Este primer enfoque se abordará mediante los modelos lineales generalizados mixtos, mientras que el segundo se analizará a través de la teoría de procesos puntuales. Finalmente, esta etapa permitirá comparar el patrón de conectividad a escala de todo el cerebro a través de técnicas de análisis multivariante.</p>

Doña María Isabel Borrajo García, Profesora Contratada Doctora de la Universidad de Santiago de Compostela y Doña Alba Vieites Prado, investigadora Ramón y Cajal del CiMUS, informan que el Trabajo Fin de Máster titulado

Aplicación de herramientas estadísticas para el análisis de la plasticidad cerebral adulta

fue realizado bajo su dirección por don Adrián Varela Rodríguez para el Máster en Técnicas Estadísticas. Estimando que el trabajo está terminado, dan su conformidad para su presentación y defensa ante un tribunal. Además, Doña María Isabel Borrajo García, Doña Alba Vieites Prado y don Adrián Varela Rodríguez

sí no

autorizan a la publicación de la memoria en el repositorio de acceso público asociado al Máster en Técnicas Estadísticas.

En Santiago de Compostela, a 3 de junio de 2026.

La tutora académica:
Doña María Isabel Borrajo García

El autor:
don Adrián Varela Rodríguez

La tutora de empresa:
Doña Alba Vieites Prado

Declaración responsable. Para dar cumplimiento a la Ley 3/2022, de 24 de febrero, de convivencia universitaria, referente al plagio en el Trabajo Fin de Máster (Artículo 11, [Disposición 2978 del BOE núm. 48 de 2022](#)), **el/la autor/a declara** que el Trabajo Fin de Máster presentado es un documento original en el que se han tenido en cuenta las siguientes consideraciones relativas al uso de material de apoyo desarrollado por otros/as autores/as:

- Todas las fuentes usadas para la elaboración de este trabajo han sido citadas convenientemente (libros, artículos, apuntes de profesorado, páginas web, programas, . . .)
- Cualquier contenido copiado o traducido textualmente se ha puesto entre comillas, citando su procedencia.
- Se ha hecho constar explícitamente cuando un capítulo, sección, demostración, . . . sea una adaptación casi literal de alguna fuente existente.

Y, acepta que, si se demostrara lo contrario, se le apliquen las medidas disciplinarias que correspondan.

Índice general

Resumen	IX
1. Introducción	1
1.1. El sistema nervioso y la plasticidad cerebral	1
1.2. CiMUS y exposición del problema	5
1.3. Descripción y objetivos del experimento	5
1.4. Etapas del experimento	7
1.4.1. Fase I: validación de los datos de partida	8
1.4.2. Fase II: comparación del patrón de conectividad a escala de todo el cerebro	9
1.5. Estructura de la memoria	9
2. Fundamentos teóricos	11
2.1. Modelos lineales generalizados mixtos con variable respuesta de conteo	11
2.1.1. Contexto y motivación de los GLMM	11
2.1.2. GLMM con variable respuesta de Poisson	13
2.1.3. GLMM con variable respuesta Binomial Negativa	17
2.2. Procesos puntuales	19
2.2.1. Conceptos previos	20
2.2.2. Comparación entre las densidades de localización de eventos de dos patrones espaciales	22
2.3. Técnicas de Análisis Multivariante	25
2.3.1. Análisis de Componentes Principales	25
2.3.2. Comparación multivariante de 2 grupos: test T^2 de Hotelling	29
3. Resultados y discusión	33
3.1. Preprocesado y preparación de los datos	33
3.2. Fase Ia: comparación entre capas de la región <i>barrel field</i>	34
3.2.1. Modelos GLMM con intercepto aleatorio para el conteo de neuronas	35
3.2.2. Modelos GLMM con intercepto y pendiente aleatorios para el conteo de neuronas	38
3.2.3. Validación y predicciones del modelo de Poisson con intercepto y pendiente aleatorios	45
3.3. Fase Ib: comparación de la distribución espacial celular	47
3.4. Fase II: comparación multivariante celular de los grupos experimentales	51
3.4.1. Aplicación del análisis de componentes principales	51
3.4.2. Aplicación del Test T^2 de Hotelling	57

4. Conclusiones y líneas futuras	59
4.1. Conclusiones	59
4.2. Líneas futuras	60
A. Bases de datos del experimento	61
B. Tablas relativas al PCA	65
Bibliografía	73

Resumen

Resumen

La plasticidad cerebral es el conjunto de mecanismos celulares y moleculares que permiten al cerebro adaptarse a cambios en el entorno interno y externo. En la edad adulta, estos mecanismos son limitados y poco conocidos, por lo que su estudio es clave para mejorar el tratamiento de alteraciones del neurodesarrollo o favorecer la recuperación tras lesiones.

Por ende, desde el laboratorio *Brain Plasticity* del CIMUS se propone analizar la plasticidad cerebral adulta a través de un modelo experimental en ratones, evaluando si el cerebro es capaz de reorganizarse y adaptarse tras la eliminación de los bigotes. En concreto, se realiza una inyección viral en la región conocida como *barrel field*, principal centro de procesamiento táctil de los bigotes, con el objetivo de analizar si dicha área es capaz de procesar información proveniente de otras modalidades sensoriales.

Para ello, es esencial comprobar la comparabilidad de las inyecciones entre los distintos animales, evaluando tanto la distribución del número de células por capas como a nivel espacial, lo que se aborda mediante los modelos lineales generalizados mixtos y la teoría de procesos puntuales, respectivamente. Finalmente, se compara el patrón de conectividad a escala de todo el cerebro mediante técnicas de análisis multivariante.

Abstract

Brain plasticity is the set of cellular and molecular mechanisms that allow the brain to adapt to changes in the internal and external environment. In adulthood, these mechanisms are limited and not well understood, so their study is key to improving the treatment of neurodevelopmental disorders or promoting recovery after injury.

Therefore, the *Brain Plasticity* laboratory at CIMUS proposes to study adult brain plasticity using an experimental mouse model, assessing whether the brain is capable of reorganizing and adapting after the removal of whiskers. In particular, a viral injection is performed in the region known as the *barrel field*, the main center for tactile processing of whiskers input, with the aim of determining whether this area is capable of processing information from other sensory modalities.

To this end, it is essential to assess the comparability of the injections across animals by evaluating both the distribution of cell numbers across layers and their spatial distribution, using generalized linear mixed models and point process theory. Finally, it is possible to compare brain-wide connectivity patterns using multivariate analysis techniques.

Capítulo 1

Introducción

Este primer capítulo tiene como finalidad contextualizar y exponer el problema abordado en el presente trabajo, esto es, el análisis de la plasticidad cerebral adulta, y se estructura como sigue. En la Sección 1.1 se establece el contexto neurobiológico del trabajo, abordando los conceptos básicos del sistema nervioso, junto con la organización de la corteza cerebral y, en especial, la plasticidad cerebral, aspectos fundamentales para la comprensión del resto del proyecto. Seguidamente, en la Sección 1.2 se detalla el problema objeto de estudio y se presenta brevemente la entidad colaboradora implicada en la realización del mismo. En este contexto, las Secciones 1.3 y 1.4 describen el diseño experimental y la metodología estadística empleada para dar respuesta a los objetivos del estudio. Finalmente, la Sección 1.5 proporciona una visión general de la estructura del proyecto, actuando como guía de lectura.

1.1. El sistema nervioso y la plasticidad cerebral

La Neurociencia es un campo multidisciplinar dedicado al estudio del **sistema nervioso**, una compleja red biológica responsable de la recepción, procesamiento y transmisión de información en el organismo. Este campo se fue consolidando a lo largo del siglo XX gracias a la integración de algunas disciplinas como la anatomía, la fisiología, la psicología y la medicina, y junto con el desarrollo de técnicas avanzadas como la electrofisiología y la neuroimagen (Sullivan et al., 2015; Yen et al., 2023).

Desde el punto de vista estructural, el sistema nervioso se divide en dos componentes: por un lado, el sistema nervioso central (SNC), compuesto por el encéfalo (que incluye el cerebro, el cerebelo y el tronco encefálico) y la médula espinal; y por otro, el sistema nervioso periférico (SNP), formado por redes de células nerviosas, denominadas neuronas, que se extienden por todo el cuerpo y conectan el sistema central con los diferentes órganos y tejidos a través de señales químicas y eléctricas (Bazira, 2021; Bach et al., 2025). Así, el SNC se encarga de procesar los estímulos y coordinar las respuestas, mientras que el SNP transmite la información sensorial desde la periferia hacia el SNC, y conduce las señales motoras desde este hacia los músculos y órganos efectores. En la Figura 1.1, obtenida de Bach et al., 2025, se ilustra de forma esquemática el funcionamiento del sistema nervioso.

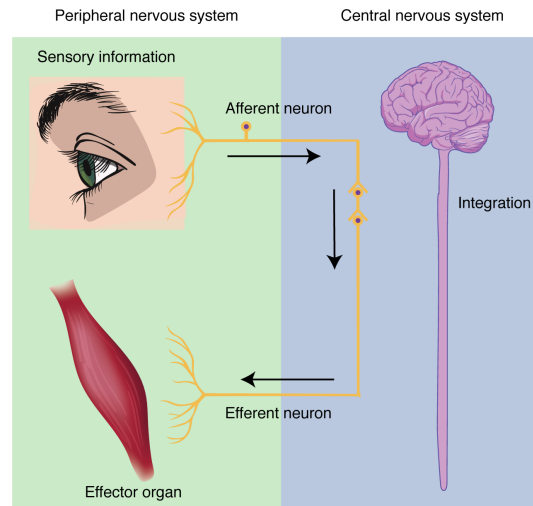


Figura 1.1. Representación del funcionamiento del sistema nervioso: el SNC recibe, procesa e interpreta la información sensorial, y envía órdenes a los órganos efectores para producir una respuesta.

En particular, desde el punto de vista morfológico, las neuronas están formadas principalmente por el soma o cuerpo celular, que contiene el núcleo y la mayor parte de los orgánulos celulares; las dendritas, encargadas de recibir información procedente de otras neuronas; y el axón, una prolongación especializada en la conducción del impulso nervioso hacia otras células. En la Figura 1.2, extraída de Bekolay, 2011, se representan las principales partes de la neurona.

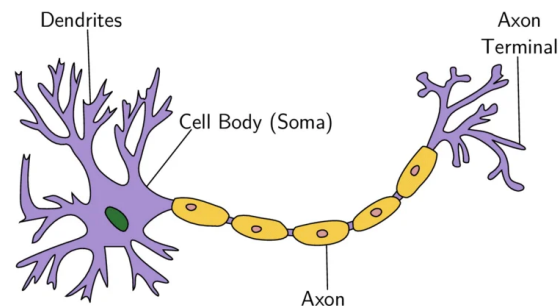


Figura 1.2. Representación esquemática de los principales componentes de la neurona.

Por otro lado, la complejidad del sistema nervioso se evidencia principalmente en el cerebro, donde la corteza cerebral¹ actúa como el principal centro de procesamiento de información. Esta región está organizada en seis capas, diferenciadas por sus tipos de neuronas y conexiones sinápticas², lo que permite la integración y procesamiento eficiente de la información neuronal (Larkum et al., 2018). Seguidamente, se describen brevemente las principales características de cada capa:

- La **Capa I**, también denominada capa molecular, es la más superficial y se caracte-

¹La corteza cerebral es la capa más externa del cerebro caracterizada por estar formada principalmente por sustancia gris, una componente esencial del SNC constituida por cuerpos neuronales y que juega un papel crucial en el procesamiento y la regulación de la información sensorial y motora.

²La sinapsis es el proceso de comunicación entre neuronas, mediante el cual se transmite información a través de señales químicas o eléctricas.

riza por una baja densidad neuronal. Está compuesta principalmente por dendritas apicales³ de neuronas piramidales⁴ que provienen de capas más profundas, así como por axones procedentes de otras áreas corticales y del tálamo⁵. Su función principal es integrar señales sinápticas procedentes de distintas neuronas, lo que permite regular la actividad de las neuronas corticales en función de la información recibida.

- La **Capa II**, conocida como capa granular externa, está formada por pequeñas neuronas piramidales y numerosas neuronas estrelladas⁶.
- La **Capa III**, denominada capa piramidal externa, contiene principalmente neuronas piramidales de tamaño medio. Esta capa, junto con la anterior, proporciona la mayor parte de proyecciones hacia otras regiones corticales, por lo que es habitual considerarlas de manera conjunta bajo la denominación Capa II/III.
- La **Capa IV**, o capa granular interna, se caracteriza por una elevada densidad de neuronas granulares⁷ y estrelladas. Esta capa constituye el principal punto de entrada de la información sensorial procedente del tálamo, especialmente en las cortezas sensoriales primarias⁸. Tras recibir esta información, las neuronas de la capa IV proyectan principalmente hacia las capas II y III, donde se inicia un procesamiento cortical más complejo.
- La **Capa V**, denominada capa piramidal interna, contiene grandes neuronas piramidales que constituyen una de las principales vías de salida de la corteza cerebral hacia estructuras subcorticales⁹. Estas neuronas proyectan hacia regiones como la médula espinal, el tronco encefálico y diferentes núcleos subcorticales. En la corteza motora primaria¹⁰, se encuentran, por ejemplo, las denominadas células de Betz, que participan en el control directo del movimiento voluntario.
- La **Capa VI**, conocida como capa multiforme, está formada por neuronas piramidales de diversos tamaños y morfologías y que proyectan hacia el tálamo. Además, se divide en dos subcapas funcionales, VIa y VIb, las cuales presentan diferencias en su organización y densidad celular. En particular, la subcapa VIa presenta una mayor densidad neuronal que la subcapa VIb.

A lo largo del trabajo, y con el objetivo de simplificar la notación, se adoptará la nomenclatura L_1 , $L_{2/3}$, L_4 , L_5 , L_{6a} y L_{6b} para referirse a las capas corticales.

³Las dendritas apicales son prolongaciones largas que surgen de la parte superior del soma de ciertos tipos de neuronas.

⁴Las neuronas piramidales son un tipo de neuronas del SNC de tamaño variable, caracterizadas por poseer un soma piramidal y que participan en la transmisión de información y en la activación de otras neuronas corticales.

⁵El tálamo es una estructura que se encuentra en el centro del encéfalo y que actúa como estación de relevo de la información sensorial y motora hacia la corteza cerebral.

⁶Las neuronas estrelladas son un tipo de neuronas del SNC caracterizadas por poseer un soma con forma de estrella especializadas en la recepción e integración de información sináptica.

⁷Las neuronas granulares son el tipo de neuronas más pequeño y numeroso del SNC.

⁸Las cortezas sensoriales primarias son las regiones de la corteza cerebral responsables de recibir e interpretar la información sensorial del organismo, es decir, aquella procedente de los diferentes sentidos (tacto, vista, oído, gusto y olfato).

⁹Las estructuras subcorticales son las regiones del encéfalo situadas por debajo de la corteza cerebral implicadas en el control motor, el procesamiento sensorial y la regulación emocional.

¹⁰La corteza motora primaria es la región de la corteza cerebral encargada de iniciar y controlar los movimientos voluntarios del organismo.

En la Figura 1.3, extraída de [Splittgerber, 2019](#), se ilustra la estructura de la corteza cerebral y su organización por capas.

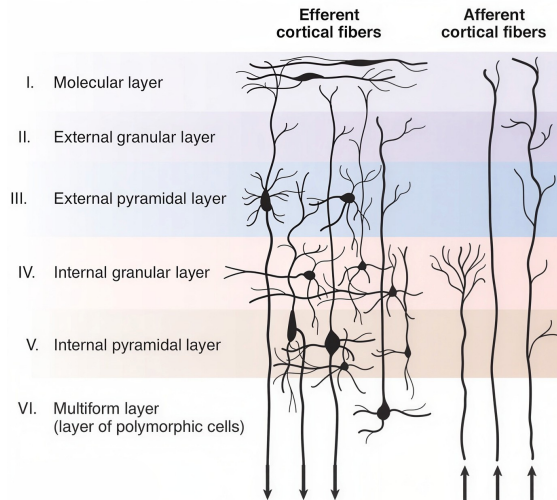


Figura 1.3. Capas de la corteza cerebral, organizadas en seis niveles con distintas estructuras y funciones neuronales.

Una de las aportaciones más relevantes de la Neurociencia radica en la comprensión de la **plasticidad cerebral**, es decir, la capacidad del cerebro de modificar su estructura y su funcionamiento en respuesta a estímulos internos o externos. Este proceso incluye cambios en las conexiones sinápticas, en la reorganización de las redes neuronales e incluso, la generación de nuevas neuronas ([Fuchsl et al., 2014](#); [Mowery et al., 2023](#)).

Aunque históricamente se consideraba que la plasticidad cerebral se limitaba solo a la infancia, en la actualidad se ha demostrado que el cerebro adulto conserva una notable capacidad de cambio y reorganización. Si bien es cierto que dicha capacidad disminuye progresivamente con la edad, esta no desaparece por completo, permaneciendo activa incluso en edades avanzadas. Esto resulta esencial para el aprendizaje continuo y la adquisición de nuevas habilidades, así como para la adaptación a cambios en el entorno o en las condiciones de vida. Asimismo, desempeña un papel clave en la recuperación funcional tras lesiones cerebrales, como ictus o traumatismos y contribuye a compensar el deterioro cognitivo asociado al envejecimiento.

Dado que la plasticidad cerebral depende de la organización y modificación de las conexiones neuronales, resulta esencial estudiar también la conectividad para comprender los mecanismos de reorganización y adaptación del cerebro. Estas conexiones se establecen a través de los axones, caracterizadas por ser extremadamente finas (< 100 nm de diámetro) y largas (desde unas micras hasta más de 1 metro en el ser humano).

Tradicionalmente, la conectividad cerebral se ha estudiado mediante diferentes técnicas que permiten visualizar los axones a diferentes escalas. Por un lado, se emplean métodos indirectos, con baja resolución pero aplicables a escala cerebral completa, como las técnicas de resonancia magnética, que incluye a la resonancia magnética funcional, orientada a estimar la relación funcional entre grandes áreas cerebrales, y a la imagen por tensor de difusión, que permite estimar la direccionalidad de los grandes tractos de fibras cere-

brales¹¹ (Lv et al., 2010). Por otro lado, existen métodos de observación directa, con alta resolución pero aplicables a pequeñas áreas cerebrales (histología en secciones o microscopía electrónica en serie), lo que limita su aplicabilidad al estudio de redes cerebrales completas (Denk et al., 2004).

Recientemente, las técnicas de clarificación tisular¹² han abierto nuevas posibilidades, permitiendo analizar la conectividad del cerebro completo con resolución subcelular. En particular, la técnica iDISCO permite transparentar los tejidos y etiquetar células o circuitos específicos, generando imágenes volumétricas¹³ que pueden cuantificarse posteriormente mediante herramientas bioinformáticas (Renier et al., 2014). Así, es posible estudiar cómo se organizan, se interconectan y se reorganizan los circuitos neuronales de todo el cerebro. En este contexto, el **Centro de Investigación en Medicina Molecular y Enfermedades Crónicas (CiMUS)** se presenta como un entorno idóneo para el desarrollo y aplicación de estas técnicas, posibilitando, entre otras cosas, el estudio de la plasticidad cerebral adulta. Seguidamente, se ofrece una breve descripción de esta entidad colaboradora y del problema planteado en el seno de la misma.

1.2. CiMUS y exposición del problema

El CiMUS es un centro de investigación biomédica con sede en Santiago de Compostela, orientado al estudio de los mecanismos moleculares implicados en diversas enfermedades crónicas, y con el objetivo de mejorar su diagnóstico, prevención y tratamiento. En este sentido, es una entidad que desarrolla diversas líneas de investigación en áreas como la neurociencia, la oncología y las enfermedades metabólicas, entre otras, y además integra enfoques multidisciplinares que favorecen el avance del conocimiento biomédico.

En particular, el laboratorio de neurociencia del CiMUS *Brain Plasticity* se dedica al estudio de la plasticidad cerebral adulta, marco en el que se desarrolla el presente Trabajo Fin de Máster. El principal objetivo de este trabajo consiste en analizar los mecanismos de la plasticidad cerebral en la etapa adulta a través de un modelo experimental en ratones, evaluando si el cerebro es capaz de reorganizarse y adaptarse tras la eliminación de las vibras faciales (bigotes), principal fuente de información somatosensorial¹⁴ en estos animales.

1.3. Descripción y objetivos del experimento

En esta sección se describe de manera detallada el diseño experimental del trabajo, en el que se emplea el ratón como modelo animal para investigar la capacidad de adaptación del cerebro adulto ante cambios permanentes en la percepción del entorno. Dado que se trata de un estudio con animales de experimentación, su desarrollo se rige por la normativa vigente en materia de bienestar animal, recogida en el [Real Decreto 53/2013](#), del 1 de febrero, en la [Directiva 2010/63/UE](#), del 22 de septiembre y en la [Orden ECC/566/2015](#),

¹¹Los tractos de fibras cerebrales son conjuntos de axones (fibras nerviosas) que conectan distintas regiones del cerebro y permiten la transmisión de información entre ellas.

¹²La clarificación tisular es una técnica de laboratorio empleada para hacer que los tejidos biológicos se vuelvan transparentes, permitiendo observar su estructura interna sin necesidad de cortarlos.

¹³Una imagen volumétrica es una representación tridimensional de un objeto o estructura anatómica, que contiene información en las tres dimensiones espaciales (alto, ancho y profundidad).

¹⁴La información somatosensorial es el conjunto de señales sensoriales procedentes del cuerpo relacionadas con el tacto, la presión, la temperatura, el dolor y la propiocepción.

del 20 de marzo. En consecuencia, las exigencias éticas establecidas por esta normativa, junto con las limitaciones económicas asociadas a este tipo de experimentos, justifican el uso de un número reducido de animales. Esto puede comprometer la robustez estadística de los resultados, aunque proporcionen una base útil para el análisis y el diseño de futuros experimentos.

En particular, se emplearon nueve ratones, divididos en dos grupos experimentales: cinco del grupo control con vibrisas faciales intactas (etiquetados como 81, 82, 95, 96 y 100) y cuatro del grupo privado sometidos a la eliminación permanente de la sensibilidad en el lado derecho del sistema de vibrisas a los 2 meses de edad (etiquetados como 104, 105, 107 y 114). Así, en este último grupo se extirpó quirúrgicamente la almohadilla de los bigotes del lado derecho de la cara, dando lugar a un modelo de privación sensorial unilateral derecho. Este modelo supone un reto adaptativo para este grupo de roedores, que deben aprender, ya en la edad adulta, a explorar su entorno con un órgano sensorial que ha pasado de ser bilateral a unilateral (funcionalmente similar a lo que implicaría la pérdida de sensibilidad en un brazo de un ser humano).

Para estudiar los posibles cambios que esta modificación induce en la conectividad entre áreas cerebrales, se emplean trazadores virales en combinación con técnicas de clarificación cerebral (iDISCO). Los trazadores virales consisten en el uso de virus adenoasociados¹⁵ (AAV, por sus siglas en inglés) modificados en el laboratorio para perder su capacidad virulenta o replicativa. A estos virus se les introduce secuencias genéticas que les permite producir proteínas fluorescentes, lo que hace posible identificar y localizar las células que infectan. Además, estos virus están diseñados para infectar subpoblaciones específicas de neuronas y, dentro de estas, según el tipo de virus, producirán marcaje anterógrado (desde el soma hasta el final de sus axones) o retrógrado (desde las terminales axónicas hacia los somas neuronales). De este modo, el uso de vectores virales permitirá “trazar” la conectividad de salida (eferencias) y de entrada (aferencias) de un área cerebral concreta en la que se inyectan los virus. En la Figura 1.4, extraída de [Haines et al., 2006](#), se ilustra el flujo axonal bidireccional de las neuronas.

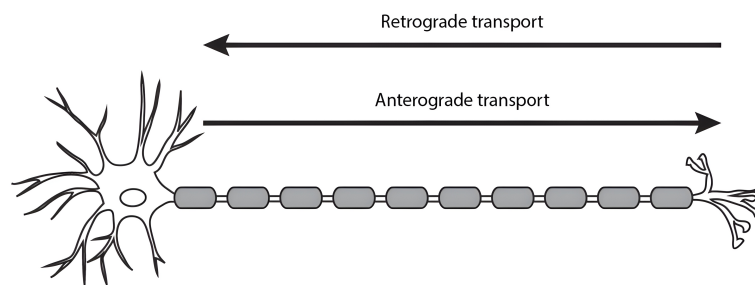


Figura 1.4. Esquema del transporte axonal bidireccional de las neuronas.

Concretamente, en este experimento se emplearon únicamente trazadores virales retrógrados dirigidos a neuronas excitatorias¹⁶, visualizados gracias a la expresión de proteína fluorescente roja (RFP). Así, cuando los roedores alcanzaron los 7 meses de edad, estos virus fueron inyectados conjuntamente en la región *barrel field* del hemisferio derecho, el

¹⁵Los virus adenoasociados son virus muy simples y no autónomos empleados como vectores para la transferencia de material genético en células.

¹⁶Las neuronas excitatorias son un tipo de células nerviosas que incrementan la probabilidad de respuesta de otras neuronas, provocando la activación del sistema nervioso.

área de la corteza cerebral encargada principalmente de procesar la información táctil de las vibrisas faciales en estos animales. En ausencia de vibrisas, el *barrel field* queda inicialmente “silenciado” (dejan de llegar impulsos con información sensorial del exterior), pero, con el paso del tiempo, existen evidencias de que esta región se “reconecta” con otras áreas cerebrales y recupera su actividad.

Tras la inyección de los virus, los animales fueron sacrificados y procesados mediante la técnica iDISCO, lo que permitió obtener imágenes de microscopía de todo el cerebro. Posteriormente, se utilizó el *software* ClearMap para la segmentación o identificación de las células rojas en el tejido cerebral. Además, ClearMap proporciona las coordenadas espaciales de cada célula, a partir de las cuales se determina la región anatómica en la que se encuentran.

Por tanto, el objetivo de este experimento es conocer qué áreas cerebrales contribuyen a la reactivación del *barrel field* localizado en el hemisferio derecho, y determinar si esta región podría procesar información de otras modalidades sensoriales (por ejemplo, auditiva o visual) o asumir un procesamiento más complejo de la información táctil procedente del otro lado de la cara (por ejemplo, reforzando su conectividad con el *barrel field* del otro hemisferio o con áreas táctiles asociativas). Seguidamente, se describen las distintas etapas del experimento.

1.4. Etapas del experimento

Tal y como se mencionó en la Sección 1.3, para el presente estudio se emplearán un total de 9 ratones (cuatro de ellos sin vibrisas faciales), tamaño muestral limitado por las consideraciones éticas y económicas. En particular, se llevará a cabo una inyección viral en la región *barrel field* del hemisferio derecho de estos animales, localizada en la corteza somatosensorial (de aproximadamente 1 mm de espesor). En la Figura 1.5, tomada de Guo et al., 2017, se representa la zona de inyección.

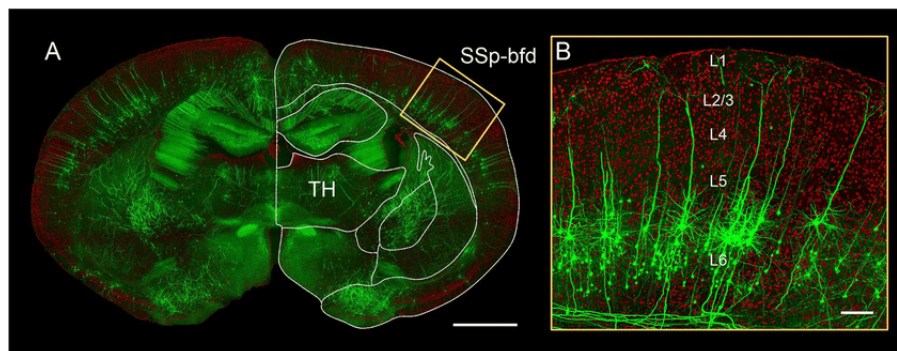


Figura 1.5. (A) Visualización del cerebro completo de un ratón obtenido mediante tomografía de dos fotones en serie (STPT). SSp-bfd denota la región *barrel field* localizada en la corteza somatosensorial primaria (zona de inyección), mientras que TH representa el tálamo. (B) Ampliación de la región *barrel field* que permite apreciar la organización laminar de la corteza cerebral, así como el marcaje anterógrado (color verde) y retrógrado (color rojo).

No obstante, este procedimiento requiere una elevada precisión, ya que se administran 200 nL de virus a unos 400 μm de profundidad, de modo que pequeñas variaciones en la

localización o en el volumen inyectado podrían afectar a la interpretación de los resultados.

Por ello, en la primera fase (Sección 1.4.1) se llevará a cabo la validación de los datos desde dos enfoques complementarios, con el objetivo de verificar la comparabilidad de las inyecciones en los distintos animales, y así garantizar la consistencia del procedimiento experimental. Una vez verificada esta condición se procederá a la segunda fase del experimento (Sección 1.4.2), en la que se comparará el patrón de conectividad a escala de todo el cerebro.

1.4.1. Fase I: validación de los datos de partida

En esta etapa, se comprobará si las inyecciones de los ratones son homogéneas desde dos perspectivas: la primera enfocada en la distribución del número de neuronas por capas del *barrel field* y la segunda en la distribución espacial de dichas células dentro de esta región.

a) Análisis de la distribución del número de células por capas del *barrel field* entre los diferentes ratones

Por un lado, pequeñas variaciones en la localización de la inyección pueden alterar el patrón de conectividad observado, ya que la corteza cerebral presenta una estructura laminar compuesta por 6 capas. Así, mientras que las neuronas de las capas superiores se conectan fuertemente con otras áreas corticales, las de las capas inferiores lo hacen principalmente con las regiones subcorticales. Por ello, es fundamental asegurar que la inyección se realice al mismo nivel entre los diferentes animales, afectando, de este modo, a neuronas de capas equivalentes. De lo contrario, se partiría de situaciones en las que la conectividad es distinta por las variaciones en la inyección, y no por la manipulación experimental (deprivación de los bigotes).

Para ello, se construirá un modelo de regresión que permita modelizar el número de neuronas (variable respuesta) según la capa cortical (variable explicativa). Como la variable de interés es de conteo, un punto natural de partida sería ajustar un modelo lineal generalizado (GLM, por sus siglas en inglés) de Poisson o, en su defecto, Binomial Negativo. No obstante, dado que cada animal es evaluado en cinco capas diferentes, el diseño presenta una estructura de medidas repetidas, lo que genera dependencia entre las observaciones dentro del mismo individuo e invalida el supuesto de independencia de los GLM. Por este motivo, se recurrirá a los modelos lineales generalizados mixtos (GLMM, por sus siglas en inglés), con distribución de Poisson o Binomial Negativa, y la elección de uno u otro se discutirá en función de la posible presencia de sobredispersión en los datos (véase Sección 2.1).

b) Análisis de la distribución espacial de las neuronas del *barrel field* entre los diferentes ratones

Por otro lado, pequeñas variaciones en el volumen de virus inyectado pueden modificar el número de células infectadas, aunque no necesariamente el patrón de conectividad, que depende de la distribución espacial de las neuronas, es decir, de cómo se organizan y se localizan dentro del *barrel field*. Por ello, es fundamental asegurar que la distribución espacial entre los animales sea similar.

Dado que las neuronas de la región *barrel field* pueden interpretarse como eventos que ocurren de manera aleatoria e independiente en el espacio tridimensional, se podrán modelar mediante los denominados procesos puntuales de Poisson (véase Sección 2.2). Asimismo, como en estos procesos la distribución espacial está caracterizada por la función de intensidad, se aplicarán diferentes métodos para comparar dicha distribución entre los nueve ratones y descartar así posibles diferencias en el marcaje celular.

1.4.2. Fase II: comparación del patrón de conectividad a escala de todo el cerebro

Tras verificar que las neuronas infectadas en la zona de inyección presenten un patrón comparable entre los animales, se procederá a comparar el patrón de conectividad entre los grupos control y privado a escala global del cerebro, considerando aquellas regiones cerebrales que contengan un número considerable de células. Dado que la zona de inyección influye en los recuentos celulares del resto de estructuras, la comparación se realizará empleando datos relativos respecto a dicha región, en lugar de recuentos absolutos.

Para ello, y dado el elevado número de regiones cerebrales, se realizará inicialmente un análisis de componentes principales, técnica multivariante que permite reducir el número de variables originales, haciendo más manejable el conjunto de datos sin perder una cantidad significativa de información. Así, a partir de este conjunto reducido, se aplicará un test de permutaciones basado en el estadístico T^2 de Hotelling para la comparación multivariante de las medias de ambos grupos.

1.5. Estructura de la memoria

El resto de la memoria se organiza en dos partes: la metodología estadística del experimento y la aplicación práctica. Así, el Capítulo 2 establece el marco metodológico, en el que se detallan las diferentes técnicas estadísticas empleadas; en particular, los modelos lineales generalizados mixtos con variable respuesta de conteo, los procesos puntuales espaciales tridimensionales y algunas técnicas de análisis multivariante. Por otro lado, en el Capítulo 3 se ofrece una descripción de las bases de datos empleadas y se recogen los resultados obtenidos en las distintas fases experimentales, junto con su correspondiente análisis y discusión. Finalmente, el Capítulo 4 sintetiza las conclusiones obtenidas, así como las limitaciones del estudio y propone líneas futuras de investigación para seguir profundizando en la comprensión de la plasticidad cerebral adulta.

Capítulo 2

Fundamentos teóricos

En este capítulo se detalla la metodología estadística que se aplicará en el Capítulo 3 al problema previamente presentado, con el fin de abordar los objetivos planteados en las dos fases del experimento. Su estructura se detalla como sigue.

En la Sección 2.1 se presentan los modelos lineales generalizados mixtos, tanto con variable respuesta de Poisson como Binomial Negativa, metodología empleada para modelizar el número de células por capas del *barrel field*. En la Sección 2.2 se presentan los procesos puntuales espaciales y se describe el contraste empleado para comparar la distribución de dos patrones espaciales, junto con su calibración tanto asintótica como *bootstrap*, lo que permite analizar la distribución espacial de las neuronas del *barrel field*. Finalmente, en la Sección 2.3 se abordan diferentes técnicas del análisis multivariante, centrándose, en particular, en el análisis de componentes principales y en el test de permutaciones basado en el estadístico T^2 de Hotelling para la comparación de medias, herramientas empleadas para estudiar las posibles diferencias en el patrón de conectividad a escala de todo el cerebro.

2.1. Modelos lineales generalizados mixtos con variable respuesta de conteo

En esta sección, se introduce, en primer lugar, el marco teórico y contextual de los **modelos lineales generalizados mixtos (GLMM)**. A partir de esta formulación, se describe el modelo de Poisson con intercepto aleatorio, planteando además la inclusión de pendiente aleatoria y de una variable de segundo nivel (véase Sección 2.1.2). Seguidamente, se estudia de forma análoga el caso en el que la variable respuesta sigue una distribución Binomial Negativa, como alternativa para abordar el problema de sobredispersión (véase Sección 2.1.3).

Las referencias fundamentales empleadas en la elaboración de esta sección serán: [Crujeiras y Conde, 2024](#), [McCullough y Nelder, 1989](#) y [Stroup et al., 2024](#).

2.1.1. Contexto y motivación de los GLMM

La regresión es una metodología estadística que permite analizar cómo una variable aleatoria de interés, Y , denominada variable respuesta, se relaciona con una o varias variables, $\mathbf{X} = (X_1, \dots, X_p)^\top$, denominadas variables explicativas. En su forma elemental, esta relación se formaliza mediante el modelo de regresión lineal, que asume que la variable

respuesta varía de modo lineal con respecto a las variables explicativas, y se expresa como sigue:

$$Y = \mathbb{E}(Y \mid \mathbf{X}) + \varepsilon = \mathbf{X}^\top \boldsymbol{\beta} + \varepsilon,$$

donde $\boldsymbol{\beta} = (\beta_1, \dots, \beta_p)^\top$ es el vector de parámetros desconocidos y ε se conoce como término de error, el cual verifica que $\mathbb{E}(\varepsilon \mid \mathbf{X} = \mathbf{x}) = 0$ para todo \mathbf{x} . Además, para que este modelo sea válido, se requiere que los errores sean independientes, homocedásticos y normalmente distribuidos.

No obstante, existen diversas situaciones en las que no se cumplen algunos de estos supuestos. Es el caso, por ejemplo, de variables de conteo, como el número de células o de variables binarias, como la presencia o ausencia de un marcador genético. Así, para abordar este tipo de variables, surgen los **modelos lineales generalizados (GLM)**, introducidos por John Nelder y Robert Wedderburn en 1972 (Nelder y Wedderburn, 1972) y posteriormente discutidos con mayor detalle en McCullough y Nelder, 1989.

Los GLM asumen que la distribución condicional de la variable respuesta pertenece a la familia exponencial, una clase de distribuciones de probabilidad que incluye a las distribuciones de Bernoulli, Poisson, Binomial Negativa o Gamma, entre otras. Así, el objetivo de este modelo es describir la relación entre la media condicional, $\mu = \mathbb{E}(Y \mid \mathbf{X})$, y el predictor lineal, $\eta = \mathbf{X}^\top \boldsymbol{\beta}$, a través de una función de enlace $g(\cdot)$, de modo que

$$g(\mu) = \eta = \mathbf{X}^\top \boldsymbol{\beta}, \quad \text{o de forma equivalente,} \quad \mu = g^{-1}(\mathbf{X}^\top \boldsymbol{\beta}).$$

Sin embargo, los GLM requieren que las observaciones sean independientes, condición que no se verifica en contextos como: datos jerárquicos, donde los individuos se organizan en grupos; medidas repetidas, donde la variable respuesta se mide más de una vez (diferentes condiciones) en cada individuo de la muestra; o datos longitudinales, que registran la evolución de la variable respuesta a lo largo del tiempo para cada sujeto. Para abordar estas limitaciones, se emplearán los denominados modelos lineales generalizados mixtos, consolidados a finales del siglo XX, y que extienden los GLM incorporando efectos aleatorios¹ en el predictor lineal, permitiendo así modelar la correlación entre observaciones dependientes (Stroup et al., 2024).

Este modelo se formula como:

$$\mathbf{y} = \mathbf{V}\boldsymbol{\beta} + \mathbf{Z}\mathbf{b},$$

donde $\mathbf{y} = (y_1, \dots, y_n)^\top$ es el vector de observaciones de la variable respuesta Y (que recordemos que pertenece a la familia exponencial), $\mathbf{V} \in \mathbb{M}_{n \times p}$ es la matriz de diseño de efectos fijos² (con p el número de efectos fijos), $\mathbf{Z} \in \mathbb{M}_{n \times q}$ es la matriz de diseño de efectos aleatorios³ (con q el número de efectos aleatorios) y $\mathbf{b} = (b_1, \dots, b_q)^\top$ es el vector de efectos aleatorios, con $\mathbf{b} \sim N(\mathbf{0}, \mathbf{G})$, siendo \mathbf{G} la matriz de varianzas-covarianzas de los efectos aleatorios.

¹Los efectos aleatorios de un modelo mixto son las componentes del modelo que representan la variabilidad entre grupos o sujetos que no puede explicarse mediante las variables explicativas (efectos fijos).

²La matriz de diseño de efectos fijos de un modelo GLMM es una representación matricial que relaciona las observaciones (filas) con las variables explicativas (columnas) del modelo.

³La matriz de diseño de efectos aleatorios de un modelo GLMM es una representación matricial que relaciona las observaciones (filas) con los efectos aleatorios (columnas) del modelo.

En particular, para la etapa de validación del experimento se dispone de una muestra de 9 ratones, y en cada individuo se obtienen cinco mediciones del número de células correspondientes a las capas corticales $L_{2/3}$, L_4 , L_5 , L_{6a} y L_{6b} , lo que constituye un diseño de medidas repetidas e introduce una estructura de dependencia intra-sujeto que debe ser adecuadamente modelada. En este contexto, se define la variable respuesta, Y_{ij} , como el número de células observadas en la medición i del ratón j , con $i \in \{1, \dots, 5\}$ y $j \in \{1, \dots, 9\}$, y la variable explicativa, X_{ij} , como un factor categórico que indica la capa cortical asociada a cada medición.

Dado que estamos ante un diseño de medidas repetidas con variable respuesta de conteo nos centraremos, desde ahora en adelante, en los modelos lineales generalizados mixtos de Poisson y Binomial Negativo. En este marco, se considerarán progresivamente distintos escenarios: en primer lugar, modelos con intercepto aleatorio, que permiten capturar diferencias en el nivel basal de la respuesta entre individuos; en segundo lugar, modelos con intercepto y pendiente aleatorios, que además incorporan variabilidad en el efecto de las covariables entre individuos; y, finalmente, modelos que incluyen una covariable de segundo nivel, es decir, una variable definida a nivel de individuo que contribuye a explicar la variabilidad interindividual.

2.1.2. GLMM con variable respuesta de Poisson

Antes de incorporar los efectos aleatorios, se analizará primero el modelo lineal generalizado de Poisson con efectos fijos. Para su formulación, resulta imprescindible especificar la distribución de la variable respuesta, que en este caso se asume que sigue una distribución de Poisson, y cuya definición formal se presenta a continuación.

Definición 2.1. *La **distribución de Poisson** es una distribución de probabilidad discreta que expresa la probabilidad de que ocurra un determinado número de eventos en un intervalo de tiempo fijo, asumiendo que los eventos suceden de forma independiente y con una tasa media de ocurrencia conocida λ . Suponiendo que la variable aleatoria Y sigue una distribución de Poisson de parámetro $\lambda > 0$, es decir, $Y \in \text{Poisson}(\lambda)$, entonces su función de probabilidad viene dada por la siguiente expresión:*

$$\mathbb{P}(Y = y) = \frac{\lambda^y e^{-\lambda}}{y!}, \quad y \in \{0, 1, 2, \dots\}.$$

Observación 2.2. *Si la variable aleatoria Y sigue una distribución de Poisson de parámetro $\lambda > 0$, entonces tanto su esperanza como su varianza coinciden y son iguales a λ :*

$$\mathbb{E}(Y) = \text{Var}(Y) = \lambda.$$

Una vez definida la distribución de la variable de interés, se procede a formular el modelo GLM de Poisson con efectos fijos.

Modelo lineal generalizado de Poisson con efectos fijos

En efecto, supongamos que la variable respuesta Y es una variable de conteo que sigue una distribución de Poisson de parámetro $\lambda > 0$. Nuestro objetivo consiste en estudiar cómo se relaciona con el conjunto de variables explicativas $\mathbf{X} = (X_1, \dots, X_p)^\top$. Para ello, construiremos un modelo para la función de regresión:

$$\lambda(\mathbf{x}) = \mathbb{E}(Y \mid \mathbf{X} = \mathbf{x}).$$

Como la variable respuesta solo puede tomar valores enteros no negativos y la función $\lambda(\mathbf{x})$ es estrictamente positiva (por ser la esperanza condicional de una variable que sigue una distribución de Poisson), se emplea la función de enlace logarítmica, $g(r) = \log(r)$, para relacionar la media condicionada con el predictor lineal $\eta = \mathbf{x}^\top \boldsymbol{\beta}$:

$$g(\lambda(\mathbf{x}, \boldsymbol{\beta})) = \log(\lambda(\mathbf{x}, \boldsymbol{\beta})) = \mathbf{x}^\top \boldsymbol{\beta},$$

donde $\mathbf{x} = (1, x_1, \dots, x_p)^\top$ y $\boldsymbol{\beta} = (\beta_0, \beta_1, \dots, \beta_p)^\top$ es el vector de parámetros. De este modo, la función de regresión del modelo de Poisson viene dada por

$$\lambda(\mathbf{x}, \boldsymbol{\beta}) = \exp\{\mathbf{x}^\top \boldsymbol{\beta}\}.$$

Una vez formulado el modelo GLM de Poisson, se planteará la inclusión de efectos aleatorios, con el fin de modelar adecuadamente la estructura de dependencia del diseño de medidas repetidas. En particular, este enfoque se desarrollará considerando únicamente una variable explicativa ($p = 1$).

Así, sea Y_{ij} la medición i en el individuo j de la variable respuesta con distribución de Poisson, comenzaremos considerando un modelo de Poisson que incluya solo un intercepto aleatorio. Este planteamiento permite capturar la variabilidad entre individuos, recogiendo posibles diferencias en el nivel basal de la variable respuesta, tal y como ocurre en el diseño experimental objeto de estudio de esta sección.

Inclusión del intercepto aleatorio

De este modo, supongamos que la variable de interés, condicionada al efecto aleatorio del intercepto individual j , b_{0j} , sigue una distribución de Poisson de parámetro λ_{ij} , es decir, $Y_{ij} | b_{0j} \sim Poisson(\lambda_{ij})$, siendo $\lambda_{ij} = \mathbb{E}(Y_{ij} | X_{ij}, b_{0j})$ la media condicional.

Bajo esta especificación, el modelo GLMM de Poisson con intercepto aleatorio viene dado por:

$$\log(\lambda_{ij}) = \beta_{0j} + \beta_1 X_{ij}, \quad j \in \{1, \dots, J\}, \quad i \in \{1, \dots, n_j\}, \quad (2.1)$$

donde $\beta_{0j} = \gamma_{00} + b_{0j}$ es el intercepto específico de cada individuo j . En particular, γ_{00} representa el intercepto promedio en la población de individuos y b_{0j} es una variable aleatoria independiente con distribución normal, es decir, $b_{0j} \sim N(0, \sigma_{b_0}^2)$.

Así, $\lambda_{ij} = \exp\{\gamma_{00} + b_{0j} + \beta_1 X_{ij}\}$ denotará la media esperada de la variable de interés para la medición i del individuo j .

No obstante, si el efecto de la variable explicativa sobre la variable respuesta no es homogéneo entre los individuos, sino que puede diferir entre ellos, será necesario añadir una pendiente aleatoria al modelo (2.1). En particular, este escenario ocurre en el diseño de medidas repetidas descrito anteriormente.

Inclusión del intercepto y pendiente aleatorios

Consideremos ahora que tanto el intercepto como la pendiente varían entre los individuos, de modo que ambos pueden modelarse como variables aleatorias a nivel individual. Para ello, supongamos que la respuesta, condicionada a los efectos aleatorios del intercepto y la pendiente del individuo j , b_{0j} y b_{1j} , sigue una distribución de Poisson de parámetro λ_{ij} ,

es decir, $Y_{ij} | b_{0j}, b_{1j} \sim \text{Poisson}(\lambda_{ij})$, siendo $\lambda_{ij} = \mathbb{E}(Y_{ij} | X_{ij}, b_{0j}, b_{1j})$ la media condicional.

Entonces, el modelo GLMM de Poisson con intercepto y pendiente aleatorios viene dado por:

$$\log(\lambda_{ij}) = \beta_{0j} + \beta_{1j}X_{ij}, \quad j \in \{1, \dots, J\}, i \in \{1, \dots, n_j\}, \quad (2.2)$$

donde $\beta_{0j} = \gamma_{00} + b_{0j}$ y $\beta_{1j} = \gamma_{10} + b_{1j}$ denotan, respectivamente, el intercepto y la pendiente aleatorios. En particular, γ_{00} y γ_{10} representan el intercepto y la pendiente promedio de la población (efectos fijos), y (b_{0j}, b_{1j}) son variables aleatorias independientes con distribución normal bivalente sobre la población de los individuos, es decir,

$$\begin{pmatrix} b_{0j} \\ b_{1j} \end{pmatrix} \sim N_2 \left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \Sigma_b = \begin{pmatrix} \sigma_{b_0}^2 & \sigma_{b_{01}} \\ \sigma_{b_{01}} & \sigma_{b_1}^2 \end{pmatrix} \right),$$

siendo $\sigma_{b_0}^2$ la varianza del intercepto, $\sigma_{b_1}^2$ la varianza de la pendiente y $\sigma_{b_{01}}$ la covarianza entre el intercepto y la pendiente.

De este modo, $\lambda_{ij} = \exp\{\gamma_{00} + b_{0j} + (\gamma_{10} + b_{1j})X_{ij}\}$ reflejará la media esperada de la respuesta para la medición i del individuo j .

Finalmente, con el fin de capturar mejor la heterogeneidad entre los individuos, se incorporará una variable explicativa de segundo nivel, W , al modelo (2.2); esto es, una característica, grupo o condición que clasifica a los sujetos de estudio y no cambia con el tiempo. Así, el nuevo modelo será capaz de detectar las posibles diferencias entre los grupos experimentales (deprivado y control), propias del diseño experimental analizado en este trabajo.

Inclusión de una variable de segundo nivel

Supongamos que la respuesta, condicionada a la variable de segundo nivel W_j y a los efectos aleatorios del individuo j , b_{0j} y b_{1j} , sigue una distribución de Poisson de parámetro λ_{ij} , es decir, $Y_{ij} | W_j, b_{0j}, b_{1j} \sim \text{Poisson}(\lambda_{ij})$, siendo $\lambda_{ij} = \mathbb{E}(Y_{ij} | X_{ij}, W_j, b_{0j}, b_{1j})$ la media condicional.

Entonces, el modelo GLMM de Poisson con intercepto y pendiente aleatorios y variable explicativa de segundo nivel viene dado por:

$$\log(\lambda_{ij}) = \beta_{0j} + \beta_{1j}X_{ij}, \quad j \in \{1, \dots, J\}, i \in \{1, \dots, n_j\}, \quad (2.3)$$

donde $\beta_{0j} = \gamma_{00} + \gamma_{01}W_j + b_{0j}$ y $\beta_{1j} = \gamma_{10} + \gamma_{11}W_j + b_{1j}$ denotan, respectivamente, el intercepto y la pendiente aleatorios. En particular, γ_{01} es el efecto de la variable de segundo nivel sobre el intercepto y γ_{11} es el coeficiente de interacción entre las variables X_{ij} y W_j .

Así, $\lambda_{ij} = \exp\{\gamma_{00} + \gamma_{01}W_j + b_{0j} + (\gamma_{10} + \gamma_{11}W_j + b_{1j})X_{ij}\}$ reflejará la media esperada de la respuesta para la medición i del individuo j .

A efectos de comprender mejor la interpretación de este modelo, se considera el siguiente ejemplo.

Ejemplo 2.3. Supóngase un estudio en el que se pretende analizar el número de intentos necesarios para resolver una tarea de memoria. Para ello, se dispone de $J = 100$ ratones y cada uno de ellos realiza la tarea durante 5 días de la semana diferentes (es decir, $i \in \{1, \dots, 5\}$). De este modo, Y_{ij} denota el número de intentos realizados por el ratón j en el día i , y constituye la variable de interés (la cual sigue una distribución de Poisson).

Con el objetivo de explicar la variabilidad de la respuesta, se considera la variable explicativa continua X_{ij} , que representa el tiempo medio de reacción del animal j en el día i , medido en milisegundos. Además, se incorpora una variable de segundo nivel W_j , que recoge una característica propia de cada animal y constante a lo largo del estudio, concretamente su pertenencia a un grupo experimental: $W_j = 1$ si el ratón pertenece a un grupo sometido a privación de sueño previa al experimento y $W_j = 0$ si pertenece al grupo control, es decir, animales que han seguido sus condiciones habituales de descanso.

De este modo, el número esperado de intentos realizados por el ratón j en el día i para resolver exitosamente la tarea de memoria viene dado por $\lambda_{ij} = \exp\{\beta_{0j} + \beta_{1j}X_{ij}\}$, donde $\beta_{0j} = \gamma_{00} + \gamma_{01}W_j + b_{0j}$ es el intercepto aleatorio y $\beta_{1j} = \gamma_{10} + \gamma_{11}W_j + b_{1j}$ es la pendiente aleatoria. Esto permite capturar la variabilidad entre los diferentes ratones, recogiendo tanto diferencias en el nivel basal de la variable respuesta, como en el efecto del tiempo de reacción sobre el número de intentos. Finalmente, nótese que si el ratón j pertenece al grupo sometido a privación de sueño, entonces su número esperado de intentos en el día i sería $\lambda_{ij} = \exp\{\gamma_{00} + \gamma_{01} + b_{0j} + (\gamma_{10} + \gamma_{11} + b_{1j})X_{ij}\}$, mientras que si pertenece al grupo control, su número esperado de intentos sería $\lambda_{ij} = \exp\{\gamma_{00} + b_{0j} + (\gamma_{10} + b_{1j})X_{ij}\}$.

Hasta este punto, se han abordado diferentes formulaciones del modelo de Poisson, desde la versión con efectos fijos, hasta el modelo con intercepto y pendiente aleatorios, y finalmente añadiendo una variable explicativa de segundo nivel que refleja la heterogeneidad entre los individuos. Sin embargo, es indispensable evaluar la hipótesis de equidispersión, un supuesto fundamental en este tipo de modelos.

Sobredispersión en el modelo de Poisson

La equidispersión es una propiedad del modelo de Poisson que exige la igualdad entre la media y la varianza de la variable respuesta. En la práctica, esta condición puede no cumplirse, dando lugar a fenómenos de sobredispersión (varianza mayor que la media) o infradispersión (varianza menor que la media), lo que limita la aplicabilidad de estos modelos.

Sea ϕ el parámetro de sobredispersión para cualquiera de las formulaciones del modelo de regresión de Poisson analizadas. Su estimación se lleva a cabo mediante la siguiente expresión:

$$\hat{\phi} = \frac{1}{n-p} \sum_{i=1}^n \left(\frac{Y_i - \hat{Y}_i}{\sqrt{\hat{Y}_i}} \right)^2 = \frac{1}{n-p} \sum_{i=1}^n \frac{(Y_i - \hat{Y}_i)^2}{\hat{Y}_i} \quad (2.4)$$

donde $\frac{Y_i - \hat{Y}_i}{\sqrt{\hat{Y}_i}}$ son los residuos de Pearson y $(n-p)$ es el número de grados de libertad residual, obtenido como la diferencia entre el número de observaciones n y el número de parámetros p del modelo. Notemos que si $\hat{\phi} > 1$, entonces hay sobredispersión, mientras que un valor $\hat{\phi} < 1$ indica infradispersión.

Como solución al problema de sobredispersión, se estudiará el modelo de regresión Binomial Negativo, considerando asimismo sus diferentes formulaciones.

2.1.3. GLMM con variable respuesta Binomial Negativa

De manera análoga al caso anterior, se analizará primero el modelo lineal generalizado Binomial Negativo con efectos fijos, para luego ampliarlo incorporando efectos aleatorios. La formulación de este modelo requiere especificar la distribución de la variable respuesta, que se asume de tipo Binomial Negativa, y cuya definición formal se presenta a continuación.

Definición 2.4. *Si se realizan repeticiones sucesivas e independientes de un ensayo de Bernoulli, con probabilidad de éxito p , entonces la variable aleatoria Y , que mide el número de éxitos obtenidos antes de que ocurra el r -ésimo fracaso, tiene una **distribución Binomial Negativa** de parámetros r y p , es decir, $Y \sim \text{BN}(r, p)$, con $r \in \mathbb{Z}^+$ y $p \in (0, 1)$. Su distribución de probabilidad viene dada por la siguiente expresión:*

$$\mathbb{P}(Y = y) = \binom{y+r-1}{y} (1-p)^r p^y, \quad y \in \{0, 1, 2, \dots\}. \quad (2.5)$$

Observación 2.5. *Si la variable aleatoria Y sigue una distribución Binomial Negativa de parámetros r y p , entonces su media y varianza vienen dadas por:*

$$\mathbb{E}(Y) = \frac{pr}{1-p} \quad y \quad \text{Var}(Y) = \frac{pr}{(1-p)^2}.$$

Notemos que la varianza es siempre mayor que la media.

En la práctica, la distribución Binomial Negativa suele reparametrizarse en términos de su media y de su parámetro de sobredispersión, lo que permite describir de forma más directa la variabilidad inherente a los datos de conteo.

En primer lugar, extenderemos el parámetro r , que originalmente tomaba valores enteros positivos, para que pueda tomar cualquier valor real positivo, es decir, $r \in (0, +\infty)$. Para ello, se sustituye el número combinatorio de la distribución de probabilidad dada en (2.5) por su expresión equivalente a través de funciones Gamma:

$$\binom{y+r-1}{y} = \frac{(y+r-1)!}{y!(r-1)!} = \frac{\Gamma(y+r)}{y!\Gamma(r)},$$

donde $\Gamma(r) = \int_0^\infty y^{r-1} e^{-y} dy$, $r \in (0, +\infty)$ es la denominada función Gamma.

A continuación, reparametrizamos la distribución Binomial Negativa en términos de su media, denotada por μ , y de su sobredispersión θ . Estos parámetros se obtienen a partir de los originales, r y p , del siguiente modo:

$$\mu = \frac{pr}{1-p} \quad y \quad \theta = r,$$

con $\mu > 0$ y $\theta > 0$. De esta forma, tendremos que $\mathbb{E}(Y) = \mu$ y $\text{Var}(Y) = \mu + \frac{\mu^2}{\theta}$, observándose que a medida que el valor de θ aumenta, la sobredispersión disminuye. Además, la función de probabilidad vendrá dada por la siguiente expresión:

$$f(y; \mu, \theta) = \mathbb{P}(Y = y) = \frac{\Gamma(y+\theta)}{y!\Gamma(\theta)} \left(\frac{\mu}{\mu+\theta}\right)^y \left(\frac{\theta}{\mu+\theta}\right)^\theta, \quad y \in \{0, 1, 2, \dots\}. \quad (2.6)$$

Tras establecer esta reparametrización, se procede a formular el modelo GLM Binomial Negativo con efectos fijos.

Modelo lineal generalizado Binomial Negativo

Supongamos que la variable respuesta Y es una variable de conteo que sigue una distribución Binomial Negativa de parámetros μ y θ . Nuestro objetivo es estudiar cómo se relaciona con el conjunto de variables explicativas $\mathbf{X} = (X_1, \dots, X_p)^\top$.

En primer lugar, notemos que $Y \sim BN(\mu(\mathbf{X}, \boldsymbol{\beta}), \theta)$, es decir, la media depende del vector de variables explicativas, mientras que el parámetro de sobredispersión θ es constante. Por ello, construiremos un modelo para la función de regresión:

$$\mu(\mathbf{x}) = \mathbb{E}(Y|\mathbf{X} = \mathbf{x}).$$

Como la variable respuesta Y solo puede tomar valores enteros no negativos y la función de regresión $\mu(\mathbf{x})$ es estrictamente positiva (por ser la esperanza condicional de una variable que sigue una distribución Binomial Negativa), entonces emplearemos la función de enlace $g(r) = \log(r)$ para relacionar la media condicionada con el predictor lineal $\eta = \mathbf{x}^\top \boldsymbol{\beta}$:

$$g(\mu(\mathbf{x}, \boldsymbol{\beta})) = \log(\mu(\mathbf{x}, \boldsymbol{\beta})) = \mathbf{x}^\top \boldsymbol{\beta},$$

donde $\mathbf{x} = (1, x_1, \dots, x_p)^\top$ y $\boldsymbol{\beta} = (\beta_0, \beta_1, \dots, \beta_p)^\top$ es el vector de parámetros. De este modo, la función de regresión de este modelo viene dada por

$$\mu(\mathbf{x}, \boldsymbol{\beta}) = \exp\{\mathbf{x}^\top \boldsymbol{\beta}\}.$$

Una vez formulado el modelo GLM Binomial Negativo, se planteará de nuevo la inclusión de efectos aleatorios, considerando únicamente una variable explicativa ($p = 1$).

Así, sea Y_{ij} la medición i en el individuo j de la variable respuesta con distribución Binomial Negativa, comenzaremos considerando un modelo Binomial Negativo que incluya solamente un intercepto aleatorio, de modo que se capture la variabilidad entre los animales del diseño experimental objeto de estudio de esta sección.

Inclusión del intercepto aleatorio

Supongamos que la variable respuesta, condicionada al efecto aleatorio del intercepto individual j , b_{0j} , sigue una distribución Binomial Negativa de parámetros μ_{ij} y θ , es decir, $Y_{ij} | b_{0j} \sim BN(\mu_{ij}, \theta)$, siendo $\mu_{ij} = \mathbb{E}(Y_{ij}|X_{ij}, b_{0j})$ la media condicional.

Entonces, el modelo GLMM Binomial Negativo con intercepto aleatorio viene dado por:

$$\log(\mu_{ij}) = \beta_{0j} + \beta_1 X_{ij}, \quad j \in \{1, \dots, J\}, \quad i \in \{1, \dots, n_j\} \quad (2.7)$$

donde $\beta_{0j} = \gamma_{00} + b_{0j}$ es el intercepto de cada individuo j .

No obstante, al igual que en el apartado anterior, puede ocurrir que el efecto de la variable explicativa sobre la variable respuesta no sea homogéneo entre los individuos, por lo que será necesario añadir una pendiente aleatoria al modelo (2.7).

Inclusión del intercepto y pendiente aleatorios

Supongamos ahora que la respuesta, condicionada a los efectos aleatorios del intercepto y la pendiente del individuo j , b_{0j} y b_{1j} , sigue una distribución Binomial Negativa de parámetros μ_{ij} y θ , es decir, $Y_{ij} | b_{0j}, b_{1j} \sim BN(\mu_{ij}, \theta)$, siendo $\mu_{ij} = \mathbb{E}(Y_{ij} | X_{ij}, b_{0j}, b_{1j})$ la media condicional.

Entonces, el modelo GLMM Binomial Negativo con intercepto y pendiente aleatorios viene dado por:

$$\log(\mu_{ij}) = \beta_{0j} + \beta_{1j}X_{ij}, \quad j \in \{1, \dots, J\}, i \in \{1, \dots, n_j\}, \quad (2.8)$$

donde $\beta_{0j} = \gamma_{00} + b_{0j}$ y $\beta_{1j} = \gamma_{10} + b_{1j}$ denotan, respectivamente, el intercepto y la pendiente aleatorios.

Finalmente, para capturar mejor la heterogeneidad entre los individuos, se incorporará al modelo (2.8) una variable explicativa de segundo nivel, W , lo que permitirá detectar las posibles diferencias entre los grupos control y deprivado del experimento.

Inclusión de una variable de segundo nivel

Supongamos que la respuesta, condicionada a la variable W_j y a los efectos aleatorios del individuo j , b_{0j} y b_{1j} , sigue una distribución Binomial Negativa de parámetros μ_{ij} y θ , es decir, $Y_{ij} | W_j, b_{0j}, b_{1j} \sim BN(\mu_{ij}, \theta)$, siendo $\mu_{ij} = \mathbb{E}(Y_{ij} | X_{ij}, W_j, b_{0j}, b_{1j})$ la media condicional.

Luego, el modelo GLMM Binomial Negativo con intercepto y pendiente aleatorios y variable explicativa de segundo nivel viene dado por:

$$\log(\mu_{ij}) = \beta_{0j} + \beta_{1j}X_{ij}, \quad j \in \{1, \dots, J\}, i \in \{1, \dots, n_j\}, \quad (2.9)$$

donde $\beta_{0j} = \gamma_{00} + \gamma_{01}W_j + b_{0j}$ y $\beta_{1j} = \gamma_{10} + \gamma_{11}W_j + b_{1j}$ denotan, respectivamente, el intercepto y la pendiente aleatorios.

2.2. Procesos puntuales

Una vez analizada la distribución laminar del número de neuronas en el *barrel field* de los distintos ratones, se procede al estudio de su distribución espacial en dicha región, como análisis complementario para validar la comparabilidad de las inyecciones entre los animales. Para ello, tal y como se comentó en el Capítulo 1, se recurre a los **procesos puntuales**.

Los procesos puntuales son una rama de la Estadística Espacial⁴ que permite modelizar la distribución aleatoria de los puntos en el espacio y con una gran aplicación en diversas disciplinas científicas. Si bien estos modelos se formulan tradicionalmente en dos dimensiones, existen muchas aplicaciones que exigen un enfoque tridimensional, como la distribución de galaxias, la localización de sismos, la ubicación de células en un tejido

⁴La Estadística Espacial es una rama de la estadística que se centra en el estudio de datos localizados en el espacio, teniendo en cuenta la dependencia entre ubicaciones cercanas. Permite estudiar patrones y distribuciones de fenómenos espaciales y se aplica en disciplinas como ecología, epidemiología y geología.

biológico... En particular, dado que en la segunda etapa del experimento se requiere analizar la distribución espacial de las neuronas cerebrales de los ratones, se ampliará la teoría de procesos puntuales al caso tridimensional.

Con el fin de analizar las posibles diferencias en la distribución neuronal de los ratones, se compararán, dos a dos, las funciones de intensidad asociadas a sus respectivos procesos espaciales (véase Sección 2.2.2). Para ello, será necesario introducir los conceptos de proceso puntual de Poisson y función de intensidad de primer orden.

Las referencias fundamentales empleadas en la elaboración de esta sección serán: [Diggle, 2013](#), [Fuentes-Santos et al., 2016](#) y [Fuentes-Santos et al., 2017](#).

2.2.1. Conceptos previos

Antes de definir en qué consiste un proceso puntual espacial, resulta pertinente presentar el concepto de proceso estocástico.

Definición 2.6. *Un **proceso estocástico** consiste en un conjunto de variables aleatorias, $\{X_d\}_{d \in D}$, definidas sobre el mismo espacio de probabilidad. El conjunto D , denominado dominio de indexación, representa los valores en los que se observa el fenómeno, pudiendo corresponder a un dominio temporal, espacial o espacio-temporal.*

En particular, a partir de un proceso estocástico con dominio espacial, se definen los procesos puntuales espaciales como sigue.

Definición 2.7. *Un **proceso puntual espacial** es un proceso estocástico que determina la localización de un número aleatorio de eventos, $\mathbf{X} = \{\mathbf{x}_1, \dots, \mathbf{x}_N\}$, distribuidos de forma irregular en una región acotada $W \subset \mathbb{R}^3$.*

Dentro de la familia de procesos puntuales espaciales, uno de los modelos más relevantes es el proceso puntual de Poisson, cuya definición se presenta a continuación.

Definición 2.8. *Un **proceso puntual de Poisson** es un proceso puntual espacial en el que los eventos ocurren de manera aleatoria e independiente dentro de una región acotada $W \subset \mathbb{R}^3$.*

En la Figura 2.1 se representa, a modo de ejemplo, una realización de un proceso puntual de Poisson definido sobre la región cúbica $W = [0, 1]^3$.

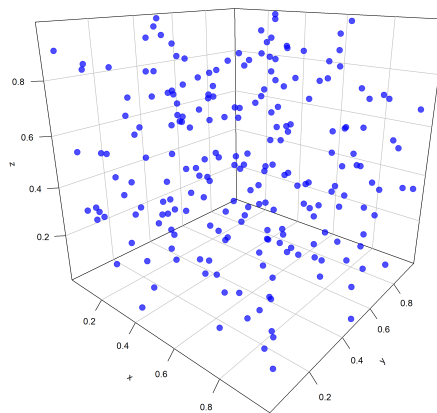


Figura 2.1. Simulación de un proceso puntual espacial de Poisson en la región $W = [0, 1]^3$ con 200 puntos.

Asimismo, para caracterizar la distribución espacial de los eventos en este tipo de procesos, se introduce el concepto de función de intensidad de primer orden.

Definición 2.9. *Dado un proceso puntual de Poisson, la distribución espacial de los eventos está determinada por la **función de intensidad de primer orden**, definida como sigue*

$$\lambda(\mathbf{x}) = \lim_{|d\mathbf{x}| \rightarrow 0} \frac{\mathbb{E}[N(d\mathbf{x})]}{|d\mathbf{x}|},$$

donde $|d\mathbf{x}|$ y $N(d\mathbf{x})$ representan, respectivamente, el volumen y el número de eventos de \mathbf{X} en la bola infinitesimal $d\mathbf{x}$ centrada en $\mathbf{x} \in W$, es decir, mide el número esperado de eventos por unidad de volumen.

Para ilustrar de manera más clara este concepto, se considera a continuación el siguiente ejemplo.

Ejemplo 2.10. Dado un ratón de laboratorio, supongamos que se selecciona una pequeña región cúbica de su tejido sanguíneo. En concreto, se considera la región $W = [0, 1]^3$, medida en mm^3 , y en donde la distribución espacial de las células se describe mediante un proceso puntual de Poisson. Además, supongamos que el comportamiento de dicha distribución puede modelarse mediante la siguiente función de intensidad de primer orden:

$$\lambda(x, y, z) = 100 \cdot \exp \left\{ -((x - 0.5)^2 + (y - 0.5)^2 + (z - 0.5)^2) \right\}.$$

De este modo, vemos que la intensidad alcanza su valor máximo en el centro del cubo, es decir, en el punto $\mathbf{p} = (0.5, 0.5, 0.5)$, mientras que decrece suavemente hacia los bordes. Luego, el mayor número esperado de células se alcanzará en una bola lo suficientemente pequeña de centro $\mathbf{p} = (0.5, 0.5, 0.5)$.

Con el fin de describir la distribución espacial relativa de los eventos, se introduce la noción de densidad de localización.

Definición 2.11. *Dado un proceso puntual de Poisson con función de intensidad de primer orden, $\lambda(\mathbf{x})$, se define la **densidad de localización de eventos** como sigue*

$$\lambda_0(\mathbf{x}) = \frac{\lambda(\mathbf{x})}{m},$$

donde $m = \int_W \lambda(\mathbf{x})$ es el número esperado de eventos del proceso puntual.

Finalmente, en función del comportamiento de la función de intensidad de primer orden, se establece la distinción entre los procesos puntuales de Poisson homogéneos e inhomogéneos.

Definición 2.12. *Un proceso puntual de Poisson es **homogéneo** si su función de intensidad de primer orden es constante, es decir,*

$$\lambda(\mathbf{x}) = \lambda > 0 \text{ para todo } \mathbf{x} \in W.$$

*En caso contrario, el proceso puntual se denomina **inhomogéneo**.*

Desde aquí en adelante, nos centraremos en los procesos puntuales de Poisson inhomogéneos.

2.2.2. Comparación entre las densidades de localización de eventos de dos patrones espaciales

Una vez presentados los conceptos elementales para poder comprender los procesos puntuales, se pretende contrastar si la densidad de localización de eventos difiere entre dos patrones espaciales dados, asumiendo que ambos son realizaciones de un proceso puntual de Poisson inhomogéneo. Para ello, se construirá un estadístico de contraste basado en la distancia L^2 , cuya calibración se obtendrá empleando tanto la aproximación normal asintótica como el método *bootstrap* suavizado.

Sea $\mathbf{X} = \{\mathbf{x}_1, \dots, \mathbf{x}_N\}$ una realización de un proceso puntual de Poisson bivariado no homogéneo en una región acotada $W \subset \mathbb{R}^3$, sean $\mathbf{X}_1 = \{\mathbf{x}_i\}_{i=1}^{N_1} = \{\mathbf{x}_{1,i}\}_{i=1}^{N_1}$ y $\mathbf{X}_2 = \{\mathbf{x}_{N_1+j}\}_{j=1}^{N_2} = \{\mathbf{x}_{2,j}\}_{j=1}^{N_2}$ los patrones espaciales de los eventos de tipo 1 y tipo 2 en \mathbf{X} , con $N = N_1 + N_2$ y denotemos por $\lambda_1(\mathbf{x})$ y $\lambda_2(\mathbf{x})$ a sus respectivas densidades de localización de eventos.

La estimación de la densidad de localización de eventos para cada patrón espacial \mathbf{X}_j , $j \in \{1, 2\}$, puede llevarse a cabo mediante la función núcleo como

$$\hat{\lambda}_{0j,H_j}(\mathbf{x}) = (e_{H_j}(\mathbf{x})N_j)^{-1} \sum_{i=1}^{N_j} k_{H_j}(\mathbf{x} - \mathbf{x}_{j,i}) \mathbb{I}[N_j \neq 0],$$

es decir,

$$\hat{\lambda}_{0j,H_j}(\mathbf{x}) = (e_{H_j}(\mathbf{x})N_j)^{-1} |H_j|^{-1/2} \sum_{i=1}^{N_j} k\left(H_j^{-1/2}(\mathbf{x} - \mathbf{x}_{j,i})\right) \mathbb{I}[N_j \neq 0],$$

donde $k(\cdot)$ es una función kernel trivariante simétrica, H_j es la matriz de ancho de banda simétrica y definida positiva, $|H_j|$ denota el determinante de H_j y

$$e_{H_j}(\mathbf{x}) = \int_W k_{H_j}(\mathbf{x} - \mathbf{y}) d\mathbf{y}$$

es el factor de corrección por efecto de borde de la región de observación.

Si los procesos puntuales \mathbf{X}_1 y \mathbf{X}_2 presentan la misma distribución espacial, entonces sus densidades de localización de eventos λ_{01} y λ_{02} coinciden. Por otro lado, condicionado al número de eventos $N_j = n_j$, los patrones espaciales observados pueden interpretarse como muestras aleatorias procedentes de distribuciones tridimensionales con densidades $\lambda_{0j}(\cdot)$, $j \in \{1, 2\}$. De este modo, es posible extender los métodos propios del análisis multivariante al contexto de los procesos puntuales espaciales.

Con el objetivo de comparar la distribución espacial de ambos procesos, se contrastará la hipótesis nula

$$H_0 : \lambda_{01}(\mathbf{x}) = \lambda_{02}(\mathbf{x}) = \lambda_0(\mathbf{x}) \text{ para casi todo } \mathbf{x} \in W. \quad (2.10)$$

Para ello, se empleará el siguiente estadístico de contraste basado en una medida de distancia L^2 :

$$\begin{aligned} \hat{T} &= \int_W \left(\hat{\lambda}_{01}(\mathbf{x}) - \hat{\lambda}_{02}(\mathbf{x}) \right)^2 d\mathbf{x} = \hat{\psi}_1 + \hat{\psi}_2 - (\hat{\psi}_{12} + \hat{\psi}_{21}) \\ &= \int_W \hat{\lambda}_{01}^2(\mathbf{x}) d\mathbf{x} + \int_W \hat{\lambda}_{02}^2(\mathbf{x}) d\mathbf{x} - \left(\int_W \hat{\lambda}_{01}(\mathbf{x}) \hat{\lambda}_{02}(\mathbf{x}) d\mathbf{x} + \int_W \hat{\lambda}_{02}(\mathbf{x}) \hat{\lambda}_{01}(\mathbf{x}) d\mathbf{x} \right), \end{aligned} \quad (2.11)$$

donde

$$\begin{aligned}\widehat{\psi}_1 &= \frac{1}{N_1^2} \sum_{i_1=1}^{N_1} \sum_{i_2=1}^{N_1} k_{H_1}(\mathbf{x}_{i_1} - \mathbf{x}_{i_2}) \mathbb{I}(N_1 > 0) \\ \widehat{\psi}_2 &= \frac{1}{N_2^2} \sum_{j_1=N_1+1}^N \sum_{j_2=N_1+1}^N k_{H_2}(\mathbf{x}_{j_1} - \mathbf{x}_{j_2}) \mathbb{I}(N_2 > 0) \\ \widehat{\psi}_{12} &= \frac{1}{N_1 N_2} \sum_{i=1}^{N_1} \sum_{j=N_1+1}^N k_{H_1}(\mathbf{x}_i - \mathbf{x}_j) \mathbb{I}(N_1 > 0) \mathbb{I}(N_2 > 0) \\ \widehat{\psi}_{21} &= \frac{1}{N_1 N_2} \sum_{i=1}^{N_1} \sum_{j=N_1+1}^N k_{H_2}(\mathbf{x}_i - \mathbf{x}_j) \mathbb{I}(N_1 > 0) \mathbb{I}(N_2 > 0),\end{aligned}$$

y donde H_1 y H_2 son las matrices de ancho de banda para los estimadores núcleo de ψ_1 y ψ_2 .

Calibración asintótica del test

El Teorema 2.13 establece que, bajo ciertas condiciones y bajo la hipótesis nula, la distribución de \widehat{T} es asintóticamente normal, lo que permite obtener una calibración asintótica del test.

Teorema 2.13. *Sea $\mathbf{X} = \{\mathbf{x}_1, \dots, \mathbf{x}_N\}$ una realización de un proceso puntual de Poisson bivariante no homogéneo en una región acotada $W \subset \mathbb{R}^3$, sean $\mathbf{X}_1 = \{\mathbf{x}_1, \dots, \mathbf{x}_{N_1}\}$ y $\mathbf{X}_2 = \{\mathbf{x}_{N_1+1}, \dots, \mathbf{x}_{N_1+N_2}\}$ los patrones espaciales de los eventos de tipo 1 y tipo 2 en \mathbf{X} con $N = N_1 + N_2$, y sean $\lambda_{01}(\mathbf{x}) = \frac{\lambda_1(\mathbf{x})}{m_1}$ y $\lambda_{02}(\mathbf{x}) = \frac{\lambda_2(\mathbf{x})}{m_2}$ sus densidades de localización de eventos respectivamente. Considerando las siguientes condiciones de regularidad para $j \in \{1, 2\}$,*

- (1) *Las densidades de localización de eventos $\lambda_{0j}(\cdot)$ tienen derivadas parciales hasta orden 2; todas sus derivadas parciales de segundo orden están acotadas, son continuas y cuadrado-integrables.*
- (2) *Las matrices de ancho de banda H_j utilizadas para calcular \widehat{T} son simétricas y definidas positivas, y tales que todas las entradas de H_j tienden a cero y*

$$m_j |H_j|^{1/2} \rightarrow \infty \quad \text{cuando } m_j \rightarrow \infty.$$

- (3) *La función núcleo $k(\cdot)$ es una función de densidad continua, simétrica y cuadrado-integrable tal que*

$$\int_{\mathbb{R}^3} \mathbf{u} \mathbf{u}^\top k(\mathbf{u}) d\mathbf{u} = \mu_2(k) I_3,$$

donde I_3 es la matriz identidad tridimensional y $0 < \mu_2(k) < \infty$.

y suponiendo que se cumple la hipótesis nula, $H_0 : \lambda_{01}(\mathbf{x}) = \lambda_{02}(\mathbf{x}) = \lambda_0(\mathbf{x})$ para casi todo $\mathbf{x} \in W$, entonces se tiene que

$$\widehat{Z} = \frac{\widehat{T} - \mu_{\widehat{T}}}{\sigma_{\widehat{T}}} \xrightarrow{d} N(0, 1), \quad (2.12)$$

donde

$$\mu_{\hat{T}} = \mathbb{E}(\hat{T}) = \left(\mathbb{E} \left[\frac{1}{N_1} \mathbb{I}(N_1 > 0) \right] |H_1|^{-1/2} + \mathbb{E} \left[\frac{1}{N_2} \mathbb{I}(N_2 > 0) \right] |H_2|^{-1/2} \right) k(\mathbf{0}),$$

y

$$\sigma_{\hat{T}}^2 = 3 \left(\mathbb{E} \left[\frac{1}{N_1} \mathbb{I}(N_1 > 0) \right] + \mathbb{E} \left[\frac{1}{N_2} \mathbb{I}(N_2 > 0) \right] \right) \left(\int \lambda_0^3(\mathbf{x}) d\mathbf{x} - \int \lambda_0^2(\mathbf{x}) d\mathbf{x} \right) (1 + o(1)).$$

Con el fin de emplear la distribución normal asintótica de \hat{T} , especificada en la expresión (2.12), es necesario estimar previamente $\mu_{\hat{T}}$ y $\sigma_{\hat{T}}$. Recordemos que los patrones espaciales observados \mathbf{X}_1 y \mathbf{X}_2 pueden interpretarse como muestras aleatorias procedentes de distribuciones aleatorias bivariantes con densidades $\lambda_{01}(\cdot)$ y $\lambda_{02}(\cdot)$ respectivamente, de modo que podemos emplear estimadores no paramétricos del marco multivariante. En particular, para $j \in \{1, 2\}$, la expresión $\mathbb{E} \left[\frac{1}{N_j} \mathbb{I}(N_j > 0) \right]$ se estima mediante $1/n_j$. De esta forma, se obtiene una calibración asintótica del test.

No obstante, esta calibración asintótica puede dar lugar a un test que no alcanza el nivel de significación, sobre todo cuando se comparan patrones puntuales con un número pequeño de eventos. Por ello, se propone emplear el método *bootstrap* suavizado para generar réplicas de \mathbf{X}_1 y \mathbf{X}_2 bajo la hipótesis nula (2.10) y así estimar la distribución nula de \hat{T} .

Calibración *bootstrap*

De acuerdo con Fuentes-Santos et al., 2017, la calibración del test mediante *bootstrap* puede implementarse de la siguiente manera:

1. Calcular el estadístico de contraste \hat{T} para los patrones espaciales \mathbf{X}_1 y \mathbf{X}_2 .
2. Sea $\mathbf{X} = \{\mathbf{X}_1, \mathbf{X}_2\}$ el patrón puntual espacial que comprende ambos tipos de eventos, obtener el estimador kernel de su densidad de localización de eventos, $\hat{\lambda}_{0,H}(\mathbf{x})$.
3. Para $b = 1, \dots, B$:

- i) Generar un proceso puntual espacial bivalente

$$\mathbf{X}_b^* = \{\mathbf{X}_{1,b}^*, \mathbf{X}_{2,b}^*\},$$

donde $\mathbf{X}_{1,b}^*$ y $\mathbf{X}_{2,b}^*$ son realizaciones de procesos puntuales de Poisson con intensidades de primer orden $n_1 \hat{\lambda}_{0,H}(\mathbf{x})$ y $n_2 \hat{\lambda}_{0,H}(\mathbf{x})$ respectivamente, siendo n_j el número de eventos en \mathbf{X}_j .

- ii) Calcular el estadístico de contraste correspondiente \hat{T}_b^* .

4. Dado un nivel de significancia α , el punto crítico es el percentil $100(1 - \alpha)$ de \hat{T}_b^* , para $b = 1, \dots, B$, es decir, $\hat{T}_{100(1-\alpha)}^*$. Así, rechazamos la hipótesis nula cuando

$$\hat{T} > \hat{T}_{100(1-\alpha)}^*.$$

Este algoritmo requiere el cálculo de la matriz de ancho de banda H en el paso 2, que puede determinarse mediante el selector de ancho de banda *plug-in* introducido por Fuentes-Santos et al., 2016. Asimismo, para los pasos 1 y 3 se necesitan determinar las matrices de ancho de banda H_1 y H_2 , que pueden obtenerse mediante el algoritmo *plug-in* propuesto por Chacón y Duong, 2010 para densidades multivariantes.

2.3. Técnicas de Análisis Multivariante

Una vez completada la validación de los datos de partida, es decir, tras comprobar que el número de neuronas infectadas en la zona de inyección sigue un patrón comparable entre individuos, se procederá a la segunda fase del experimento: la comparación del patrón de conectividad a escala global del cerebro.

Para ello, se comparará el número de neuronas entre las distintas regiones de los grupos control y privado. Dada la alta dimensionalidad de los datos, se aplicará inicialmente el análisis de componentes principales, una herramienta de análisis multivariante que permite reducir el número de variables. Así, a partir de este conjunto reducido, se empleará el test de permutaciones basado en el estadístico T^2 de Hotelling para la comparación multivariante de las medias de ambos grupos, el cual no requiere el supuesto de normalidad.

De este modo, a lo largo de esta sección se desarrollarán en detalle estas técnicas multivariantes, y cuyas referencias clave serán: [Chung y Romano, 2016](#), [Dunteman, 1989](#) y [Mason y Young, 2011](#).

2.3.1. Análisis de Componentes Principales

El **análisis de componentes principales** (PCA, por sus siglas en inglés) es una técnica estadística multivariante que permite transformar un conjunto de variables correlacionadas en un conjunto más pequeño de variables no correlacionadas denominadas componentes principales. Estas nuevas variables son combinaciones lineales de las originales y permitirán explicar gran parte de la variabilidad total de los datos. De este modo, el PCA facilitará la visualización, la interpretación y el análisis de los datos, incluso en escenarios de gran complejidad, tal y como ocurre con los datos del experimento considerado. En particular, como se mencionó anteriormente, el objetivo inicial de esta etapa será reducir la dimensionalidad del conjunto de variables regionales $\mathbf{X} = (X_1, \dots, X_{132})^\top$, donde cada variable representa el número de neuronas en una determinada estructura cerebral, relativo al número de neuronas de la zona de inyección.

Definición 2.14. Sea $\mathbf{X} = (X_1, \dots, X_p)^\top$ un vector aleatorio p -dimensional con vector de medias $\boldsymbol{\mu}$ y matriz de covarianzas $\boldsymbol{\Sigma}$. Entonces, la **primera componente principal** de \mathbf{X} se define como la variable aleatoria Y_1 , obtenida mediante una combinación lineal normalizada de las variables de \mathbf{X} y que maximiza la variabilidad, es decir:

$$Y_1 = \mathbf{a}_1^\top \mathbf{X} = a_{11}X_1 + \dots + a_{p1}X_p, \quad \text{con } \mathbf{a}_1 = (a_{11}, \dots, a_{p1})^\top \in \mathbb{R}^p$$

y

$$\text{Var}(Y_1) = \max \left\{ \text{Var}(\mathbf{a}^\top \mathbf{X}) : \mathbf{a} \in \mathbb{R}^p, \mathbf{a}^\top \mathbf{a} = 1 \right\}.$$

Una vez definida la primera componente principal, construiremos de manera secuencial las restantes componentes principales, que estarán sujetas a condiciones de ortogonalidad entre sí.

Definición 2.15. Las p **componentes principales** de \mathbf{X} se definen como las variables aleatorias (Y_1, \dots, Y_p) tales que

$$Y_j = \mathbf{a}_j^\top \mathbf{X} = a_{1j}X_1 + \dots + a_{pj}X_p, \quad \text{con } \mathbf{a}_j = (a_{1j}, \dots, a_{pj})^\top \in \mathbb{R}^p, \quad j \in \{1, \dots, p\}$$

y

$$\begin{aligned}\text{Var}(Y_1) &= \text{máx} \left\{ \text{Var}(\mathbf{a}^\top \mathbf{X}) : \mathbf{a} \in \mathbb{R}^p, \mathbf{a}^\top \mathbf{a} = 1 \right\}, \\ \text{Var}(Y_2) &= \text{máx} \left\{ \text{Var}(\mathbf{a}^\top \mathbf{X}) : \mathbf{a} \in \mathbb{R}^p, \mathbf{a}^\top \mathbf{a} = 1, \mathbf{a}^\top \boldsymbol{\Sigma} \mathbf{a}_1 = 0 \right\}, \\ \text{Var}(Y_j) &= \text{máx} \left\{ \text{Var}(\mathbf{a}^\top \mathbf{X}) : \mathbf{a} \in \mathbb{R}^p, \mathbf{a}^\top \mathbf{a} = 1, \mathbf{a}^\top \boldsymbol{\Sigma} \mathbf{a}_1, \dots, \mathbf{a}^\top \boldsymbol{\Sigma} \mathbf{a}_{j-1} = 0 \right\}, j \in \{3, \dots, p\}.\end{aligned}$$

A continuación, se describe el procedimiento mediante el cual se obtienen las componentes principales de forma sucesiva.

Extracción de las componentes principales

Dado que la varianza de $\mathbf{a}^\top \mathbf{X}$ viene dada por $\mathbf{a}^\top \boldsymbol{\Sigma} \mathbf{a}$, el cálculo de las p componentes principales puede formularse como una sucesión de problemas de optimización.

Así, la **primera componente principal** se obtiene a partir del vector \mathbf{a}_1 , el cual es solución del siguiente problema de maximización:

$$\begin{aligned}\text{máx}_{\mathbf{a}_1 \in \mathbb{R}^p} \quad & \mathbf{a}_1^\top \boldsymbol{\Sigma} \mathbf{a}_1 \\ \text{sujeto a} \quad & \mathbf{a}_1^\top \mathbf{a}_1 = 1\end{aligned}\tag{2.13}$$

Para resolverlo, se emplea el método de los *multiplicadores de Lagrange*. De este modo, se define la función lagrangiana de (2.13) como

$$\mathcal{L}(\mathbf{a}_1, \lambda) = \mathbf{a}_1^\top \boldsymbol{\Sigma} \mathbf{a}_1 - \lambda(\mathbf{a}_1^\top \mathbf{a}_1 - 1),$$

y derivando e igualando a cero, obtendremos su máximo:

$$\nabla_{\mathbf{a}_1} \mathcal{L} = 2\boldsymbol{\Sigma} \mathbf{a}_1 - 2\lambda \mathbf{I}_p \mathbf{a}_1 = (\boldsymbol{\Sigma} - \lambda \mathbf{I}_p) \mathbf{a}_1 = 0.\tag{2.14}$$

Dado que estamos ante un sistema lineal de ecuaciones, el teorema de Rouché-Frobenius garantiza que dicho sistema tendrá una solución distinta de cero si la matriz $(\boldsymbol{\Sigma} - \lambda \mathbf{I}_p)$ es singular, es decir, si $\det(\boldsymbol{\Sigma} - \lambda \mathbf{I}_p) = 0$. Luego, λ será un autovalor de la matriz $\boldsymbol{\Sigma}$.

Finalmente, de la expresión (2.14) deducimos que $\boldsymbol{\Sigma} \mathbf{a}_1 = \lambda \mathbf{I}_p \mathbf{a}_1$, por lo que

$$\text{Var}(Y_1) = \mathbf{a}_1^\top \boldsymbol{\Sigma} \mathbf{a}_1 = \mathbf{a}_1^\top \lambda \mathbf{I}_p \mathbf{a}_1 = \lambda \mathbf{a}_1^\top \mathbf{a}_1 = \lambda.$$

Por tanto, para que la varianza de la primera componente principal alcance su valor máximo, se tendrá que tomar el mayor autovalor de la matriz $\boldsymbol{\Sigma}$, denotado por λ_1 , con autovector asociado \mathbf{a}_1 . Además, notemos que por ser $\boldsymbol{\Sigma}$ una matriz semidefinida positiva dicho autovalor es necesariamente no negativo.

La **segunda componente principal** se obtiene a partir del vector \mathbf{a}_2 , resolviendo el siguiente problema:

$$\begin{aligned}\text{máx}_{\mathbf{a}_2 \in \mathbb{R}^p} \quad & \mathbf{a}_2^\top \boldsymbol{\Sigma} \mathbf{a}_2 \\ \text{sujeto a} \quad & \mathbf{a}_2^\top \mathbf{a}_2 = 1 \\ & \mathbf{a}_2^\top \boldsymbol{\Sigma} \mathbf{a}_1 = 0,\end{aligned}\tag{2.15}$$

donde la última restricción es consecuencia de la incorrelación entre las componentes principales Y_2 e Y_1 . Además, como $\Sigma \mathbf{a}_1 = \lambda \mathbf{a}_1$, entonces esta restricción es equivalente a que $\mathbf{a}_2^\top \mathbf{a}_1 = 0$, es decir, a que ambos vectores sean ortogonales.

Así, definimos la función lagrangiana de (2.15) como

$$\mathcal{L}(\mathbf{a}_2, \lambda, \mu) = \mathbf{a}_2^\top \Sigma \mathbf{a}_2 - \lambda(\mathbf{a}_2^\top \mathbf{a}_2 - 1) - \mu \mathbf{a}_2^\top \mathbf{a}_1,$$

y derivando e igualando a cero, tenemos que

$$\nabla_{\mathbf{a}_2} \mathcal{L} = 2\Sigma \mathbf{a}_2 - 2\lambda \mathbf{a}_2 - \mu \mathbf{a}_1 = 0. \quad (2.16)$$

Finalmente, multiplicando esta última expresión por el vector \mathbf{a}_1 , que recordemos que es unitario y ortogonal a \mathbf{a}_2 , obtenemos que $2\mathbf{a}_1^\top \Sigma \mathbf{a}_2 - \mu = 0$, y como Y_1 e Y_2 están incorreladas, entonces $\mu = 2\mathbf{a}_1^\top \Sigma \mathbf{a}_2 = 2\mathbf{a}_2^\top \Sigma \mathbf{a}_1 = 0$.

De este modo, la expresión (2.16) puede reescribirse como

$$(\Sigma - \lambda \mathbf{I}_p) \mathbf{a}_2 = 0,$$

y de modo análogo al caso anterior, el escalar λ se corresponde con el segundo mayor autovalor de Σ , con autovector asociado \mathbf{a}_2 . Lo denotamos por λ_2 y representa el máximo valor que alcanza la varianza de Y_2 , siendo además no negativo.

Siguiendo el mismo procedimiento, se obtienen las demás componentes principales, Y_j , $j \in \{3, \dots, p\}$, incorreladas con las componentes anteriores y con varianzas decrecientes $\text{Var}(Y_3) = \lambda_3 \geq \dots \geq \text{Var}(Y_p) = \lambda_p \geq 0$.

Descomposición de la variabilidad

Ahora, veamos como obtener la denominada descomposición espectral de la matriz Σ , expresión empleada para calcular la proporción de varianza explicada.

En efecto, recordemos que las p componentes principales de \mathbf{X} se definen como

$$Y_j = \mathbf{a}_j^\top \mathbf{X}, \quad j \in \{1, \dots, p\},$$

donde \mathbf{a}_j es el autovector asociado a λ_j de norma uno. Matricialmente, se obtiene que

$$\mathbf{Y} = \mathbf{A}\mathbf{X},$$

siendo \mathbf{A} la matriz que recoge los autovectores por columnas.

Dado que $\text{Var}(Y_j) = \lambda_j$, $j \in \{1, \dots, p\}$ y que las p componentes principales son variables incorreladas, es decir, $\text{Cov}(Y_i, Y_j) = 0$ si $i \neq j$, entonces la matriz de covarianzas de \mathbf{Y} viene dada por

$$\mathbf{\Lambda} = \begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_p \end{pmatrix}.$$

Además,

$$\mathbf{\Lambda} = \text{Var}(\mathbf{Y}) = \text{Var}(\mathbf{A}^\top \mathbf{X}) = \mathbf{A}^\top \text{Var}(\mathbf{X}) \mathbf{A} = \mathbf{A}^\top \mathbf{\Sigma} \mathbf{A},$$

y como \mathbf{A} es una matriz ortogonal, es decir, $\mathbf{A}^\top \mathbf{A} = \mathbf{A} \mathbf{A}^\top = \mathbf{I}_p$, obtenemos que

$$\mathbf{\Sigma} = \mathbf{A} \mathbf{\Lambda} \mathbf{A}^\top.$$

Esto se conoce como la descomposición espectral de la matriz $\mathbf{\Sigma}$.

Finalmente, veamos cómo cuantificar la importancia de cada componente principal a través de la proporción de varianza explicada que aporta a la variabilidad total del conjunto de datos.

Proporción de varianza explicada

En primer lugar, sumando todos los autovalores de la matriz $\mathbf{\Sigma}$, obtenemos la varianza total de las componentes:

$$\sum_{i=1}^p \text{Var}(Y_i) = \sum_{i=1}^p \lambda_i,$$

que no es más que la traza de la matriz diagonal de autovalores, $\text{tr}(\mathbf{\Lambda})$.

Así, empleando la descomposición espectral de la matriz $\mathbf{\Sigma}$, las propiedades del operador traza y que \mathbf{A} es una matriz ortogonal, entonces se verifica que

$$\text{tr}(\mathbf{\Lambda}) = \text{tr}(\mathbf{A}^\top \mathbf{\Sigma} \mathbf{A}) = \text{tr}(\mathbf{\Sigma} \mathbf{A}^\top \mathbf{A}) = \text{tr}(\mathbf{\Sigma}).$$

De este modo, obtenemos que

$$\sum_{i=1}^p \text{Var}(Y_i) = \text{tr}(\mathbf{\Lambda}) = \text{tr}(\mathbf{\Sigma}) = \sum_{i=1}^p \text{Var}(X_i),$$

es decir, la variabilidad total de las variables originales coincide con la variabilidad total de las componentes principales. Por tanto, la proporción de variabilidad explicada por la componente principal j viene dada por

$$\frac{\lambda_j}{\sum_{i=1}^p \lambda_i}, \quad j \in \{1, \dots, p\}. \quad (2.17)$$

Por otro lado, si extraemos m componentes principales, tendremos que

$$\frac{\sum_{i=1}^m \lambda_i}{\sum_{i=1}^m \lambda_i + \sum_{i=m+1}^p \lambda_i} \quad (2.18)$$

es la proporción de variabilidad explicada por las m primeras componentes principales.

En la práctica, existen varios criterios para decidir entre la simplicación debida a la reducción de la dimensionalidad y la pérdida de información de la variabilidad no explicada:

- **Criterio de la varianza explicada.** Se emplea el número de componentes principales que expliquen conjuntamente una proporción de varianza establecida, habitualmente el 90 % o 95 %.
- **Gráfico de sedimentación.** Se representan gráficamente los autovalores $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p$ en orden decreciente, con el objetivo de buscar un “codo” en el gráfico.

- **Criterio del valor propio (Kaiser).** Se seleccionan sólo las componentes cuyos autovalores sean mayores que la media, es decir, mayores que $\text{tr}(\mathbf{\Sigma})/p$.

Finalmente, conviene señalar que las componentes principales son sensibles a cambios de escala de las variables originales. Por ello, si por ejemplo se incrementa la escala de una variable, su varianza aumentará, de modo que la primera componente principal va a tender a esta variable. Por este motivo, es habitual obtener las componentes principales a partir de las variables estandarizadas, es decir, transformadas de modo que tengan media cero y varianza uno. Esta estandarización equivale a realizar el PCA sobre la matriz de correlaciones en lugar de la matriz de covarianzas.

2.3.2. Comparación multivariante de 2 grupos: test T^2 de Hotelling

Por otro lado, el **test T^2 de Hotelling** es una técnica estadística multivariante que extiende la prueba t de Student al caso en el que se analizan simultáneamente múltiples variables, permitiendo evaluar, bajo ciertas condiciones, si existen diferencias significativas entre los vectores de medias de dos poblaciones.

Supongamos que disponemos de 2 muestras aleatorias simples e independientes entre sí, cada una de ellas procedente de una población normal multivariante con matriz de covarianzas común $\mathbf{\Sigma}$ (hipótesis de homocedasticidad):

$$\begin{aligned} \mathbf{Y}_1^{(G_1)}, \dots, \mathbf{Y}_{n_1}^{(G_1)} &\sim N_p(\boldsymbol{\mu}^{(G_1)}, \mathbf{\Sigma}) \\ \mathbf{Y}_1^{(G_2)}, \dots, \mathbf{Y}_{n_2}^{(G_2)} &\sim N_p(\boldsymbol{\mu}^{(G_2)}, \mathbf{\Sigma}), \end{aligned}$$

donde $\boldsymbol{\mu}^{(G_i)} \in \mathbb{R}^p$ es el vector de medias del grupo G_i y con tamaños muestrales respectivos n_1 y n_2 .

El objetivo será contrastar la hipótesis nula de que los vectores de medias p -dimensionales sean iguales, es decir:

$$H_0 : \boldsymbol{\mu}^{(G_1)} = \boldsymbol{\mu}^{(G_2)}. \quad (2.19)$$

Para contrastar la hipótesis nula (2.19), emplearemos el test T^2 de Hotelling. Para ello, se definen, en primer lugar, los vectores de medias muestrales de los grupos G_1 y G_2 como

$$\bar{\mathbf{Y}}^{(G_1)} = \frac{1}{n_1} \sum_{j=1}^{n_1} \mathbf{Y}_j^{(G_1)} \quad \text{y} \quad \bar{\mathbf{Y}}^{(G_2)} = \frac{1}{n_2} \sum_{j=1}^{n_2} \mathbf{Y}_j^{(G_2)}.$$

Seguidamente, se determinan las matrices de covarianza muestrales de ambos grupos:

$$\hat{\mathbf{\Sigma}}_1 = \frac{1}{n_1 - 1} \sum_{j=1}^{n_1} \left(\mathbf{Y}_j^{(G_1)} - \bar{\mathbf{Y}}^{(G_1)} \right) \left(\mathbf{Y}_j^{(G_1)} - \bar{\mathbf{Y}}^{(G_1)} \right)^\top$$

y

$$\hat{\mathbf{\Sigma}}_2 = \frac{1}{n_2 - 1} \sum_{j=1}^{n_2} \left(\mathbf{Y}_j^{(G_2)} - \bar{\mathbf{Y}}^{(G_2)} \right) \left(\mathbf{Y}_j^{(G_2)} - \bar{\mathbf{Y}}^{(G_2)} \right)^\top,$$

obteniendo así la estimación insesgada de la matriz de covarianza agrupada

$$\hat{\mathbf{\Sigma}} = \frac{(n_1 - 1)\hat{\mathbf{\Sigma}}_1 + (n_2 - 1)\hat{\mathbf{\Sigma}}_2}{n_1 + n_2 - 2}.$$

De este modo, el **estadístico de contraste T^2 de Hotelling** para dos muestras viene dado por:

$$T^2 = \frac{n_1 n_2}{n_1 + n_2} \left(\bar{\mathbf{Y}}^{(G_1)} - \bar{\mathbf{Y}}^{(G_2)} \right)^\top \hat{\Sigma}^{-1} \left(\bar{\mathbf{Y}}^{(G_1)} - \bar{\mathbf{Y}}^{(G_2)} \right), \quad (2.20)$$

que bajo la hipótesis nula tiene una distribución T^2 de Hotelling de parámetros p y $n_1 + n_2 - 1$. Además, se puede relacionar con la distribución F de Snédecor a través de

$$\frac{n_1 + n_2 - p - 1}{(n_1 + n_2 - 2)p} T^2 \sim F(p, n_1 + n_2 - 1 - p).$$

No obstante, existen muchos casos en las que no se cumplen los supuestos clásicos de normalidad u homocedasticidad, lo que limita la validez del test T^2 de Hotelling. Para abordar esta situación, se propone el test de permutaciones T^2 de Hotelling, una alternativa no paramétrica que no asume supuestos sobre las poblaciones de partida.

Test de permutaciones

Supongamos que disponemos de las muestras aleatorias simples e independientes entre sí

$$\mathbf{Y}_1^{(G_1)}, \dots, \mathbf{Y}_{n_1}^{(G_1)} \quad \text{e} \quad \mathbf{Y}_1^{(G_2)}, \dots, \mathbf{Y}_{n_2}^{(G_2)},$$

donde las observaciones p -dimensionales del grupo G_1 son i.i.d. según la distribución de probabilidad P , con vector de medias $\boldsymbol{\mu}^{(G_1)}$ y matriz de covarianza $\boldsymbol{\Sigma}_1$, y las observaciones p -dimensionales del grupo G_2 son i.i.d. según la distribución Q , con vector de medias $\boldsymbol{\mu}^{(G_2)}$ y matriz de covarianza $\boldsymbol{\Sigma}_2$. Consideremos, además que, $n = n_1 + n_2$, y agrupemos las observaciones de ambos grupos como sigue:

$$\mathbf{Y} = \left(\mathbf{Y}_1^{(G_1)}, \dots, \mathbf{Y}_{n_1}^{(G_1)}, \mathbf{Y}_1^{(G_2)}, \dots, \mathbf{Y}_{n_2}^{(G_2)} \right) = (\mathbf{Y}_1, \dots, \mathbf{Y}_{n_1}, \mathbf{Y}_{n_1+1}, \dots, \mathbf{Y}_n).$$

Nuestro objetivo será contrastar la hipótesis nula (2.19), lo que se traduce en

$$\begin{cases} H_0 : \mu_k^{(G_1)} = \mu_k^{(G_2)} & \text{para todo } k \in \{1, \dots, p\}, \\ H_a : \mu_k^{(G_1)} \neq \mu_k^{(G_2)} & \text{para algún } k \in \{1, \dots, p\}. \end{cases} \quad (2.21)$$

Para ello, seleccionaremos un estadístico de contraste que sea asintóticamente libre de distribución, es decir, cuya distribución límite no dependa de la distribución original de los datos, de modo que la denominada distribución de permutaciones aproxime asintóticamente la verdadera distribución muestral incondicional del estadístico de prueba.

En primer lugar, definamos qué entendemos por distribución de permutaciones de un estadístico de contraste.

Definición 2.16. *Dado un estadístico de contraste multivariante \mathbf{T}_{n_1, n_2} para el contraste (2.21), definimos su **distribución de permutaciones** como*

$$\hat{\mathbf{R}}^{\mathbf{T}_{n_1, n_2}}(\mathbf{t}) = \frac{1}{n!} \sum_{\pi \in J_n} \mathbb{I} \left\{ \mathbf{T}_{n_1, n_2}(Y_{\pi(1)}, \dots, Y_{\pi(n)}) \leq \mathbf{t} \right\}, \quad (2.22)$$

donde J_n denota el conjunto de las $n!$ permutaciones de $\{1, 2, \dots, n\}$ y $\mathbf{t} \in \mathbb{R}^p$. Es decir, no es más que la función de distribución empírica del estadístico de contraste \mathbf{T}_{n_1, n_2} recalculado sobre todas las permutaciones de los datos.

Además, como se señaló anteriormente, para garantizar la validez asintótica de los tests de permutaciones, es fundamental emplear un estadístico de contraste asintóticamente pivotal, es decir, cuya distribución límite no dependa de las distribuciones subyacentes P y Q . En particular, se empleará el estadístico T^2 de Hotelling modificado, ya que para el estadístico de contraste clásico (2.20) se requieren los supuestos de normalidad y homocedasticidad entre grupos. Así, esta modificación permitirá alcanzar una probabilidad de rechazo asintótica igual a α , manteniendo el nivel exacto cuando las distribuciones coinciden.

En primer lugar, se define el estadístico \mathbf{T}_{n_1, n_2} , basado en la diferencia de los vectores de medias muestrales, como

$$\mathbf{T} = \begin{pmatrix} T_1 \\ \vdots \\ T_p \end{pmatrix} = \begin{pmatrix} \sqrt{n_1} \left(\frac{1}{n_1} \sum_{i=1}^{n_1} Y_{i,1}^{(G_1)} - \frac{1}{n_2} \sum_{j=1}^{n_2} Y_{j,1}^{(G_2)} \right) \\ \vdots \\ \sqrt{n_1} \left(\frac{1}{n_1} \sum_{i=1}^{n_1} Y_{i,p}^{(G_1)} - \frac{1}{n_2} \sum_{j=1}^{n_2} Y_{j,p}^{(G_2)} \right) \end{pmatrix} = \sqrt{n_1} \left[\frac{1}{n_1} \sum_{i=1}^{n_1} \mathbf{Y}_i^{(G_1)} - \frac{1}{n_2} \sum_{j=1}^{n_2} \mathbf{Y}_j^{(G_2)} \right].$$

A partir de él, se define el siguiente estadístico de contraste, del cual se deduce posteriormente la expresión del estadístico T^2 de Hotelling modificado:

$$\mathbf{S}_{n_1, n_2} = \hat{\Sigma}^{-1/2} \mathbf{T}_{n_1, n_2} = n_1^{-1/2} \hat{\Sigma}^{-1/2} \left[\sum_{i=1}^{n_1} \mathbf{Y}_i^{(G_1)} - \frac{n_1}{n_2} \sum_{j=1}^{n_2} \mathbf{Y}_j^{(G_2)} \right], \quad (2.23)$$

donde

$$\hat{\Sigma} = \hat{\Sigma}_1 + \frac{n_1}{n_2} \hat{\Sigma}_2,$$

y siendo las matrices $\hat{\Sigma}_1$ y $\hat{\Sigma}_2$ estimadores consistentes⁵ de Σ_1 y Σ_2 . De hecho, en el siguiente teorema, se establece que la distribución de permutaciones de este estadístico converge asintóticamente a una distribución normal estándar.

Teorema 2.17. *Supongamos que para cada componente $k \in \{1, \dots, p\}$,*

$$\mathbb{E} \left(Y_k^{(G_1)} \right) = \mathbb{E} \left(Y_k^{(G_2)} \right),$$

y

$$0 < \text{Var} \left(Y_{i,k}^{(G_1)} \right) < \infty \quad y \quad 0 < \text{Var} \left(Y_{j,k}^{(G_2)} \right) < \infty.$$

Sea $n_1 \rightarrow \infty$ y $n_2 \rightarrow \infty$. Además, si $q = \lim \frac{n_1}{n_1 + n_2}$, asumimos que

$$q_{n_1} = \frac{n_1}{n} \rightarrow q \in [0, 1), \quad \text{con} \quad q_{n_1} - q = O(n_1^{-1/2}).$$

Entonces, la distribución de permutaciones para el estadístico multivariante \mathbf{S}_{n_1, n_2} , $\hat{\mathbf{R}}^{\mathbf{S}_{n_1, n_2}}$, satisface que

$$\sup_{\mathbf{t} \in \mathbb{R}^p} \left| \hat{\mathbf{R}}^{\mathbf{S}_{n_1, n_2}}(\mathbf{t}) - \Phi_p(\mathbf{t}) \right| \xrightarrow{p} 0,$$

donde Φ_p denota la distribución normal estándar multivariante de dimensión p .

⁵Un estimador es consistente si, a medida que aumenta el tamaño muestral, se aproxima cada vez más al verdadero valor del parámetro poblacional.

Ahora, se define a partir de (2.23), el **estadístico de contraste T^2 de Hotelling modificado** como

$$\tilde{T}_{n_1, n_2}^2 = \|\mathbf{S}_{n_1, n_2}\|^2 = \mathbf{T}_{n_1, n_2}^\top \hat{\Sigma}^{-1} \mathbf{T}_{n_1, n_2}. \quad (2.24)$$

Nótese que este estadístico es una modificación del estadístico clásico T^2 de Hotelling (2.20), el cual puede expresarse de forma equivalente como

$$T_{n_1, n_2}^2 = \mathbf{T}_{n_1, n_2}^\top \bar{\Sigma}^{-1} \mathbf{T}_{n_1, n_2},$$

donde

$$\bar{\Sigma} = \left(\frac{n_1 + n_2}{n_2} \right) \frac{(n_1 - 1)\hat{\Sigma}_1 + (n_2 - 1)\hat{\Sigma}_2}{n_1 + n_2 - 2}.$$

Finalmente, en el siguiente teorema se establece que el límite asintótico de la distribución de permutaciones del estadístico \tilde{T}_{n_1, n_2}^2 es una distribución chi-cuadrado con p grados de libertad.

Teorema 2.18. *Bajo las condiciones del Teorema 2.17, consideremos el estadístico T^2 de Hotelling modificado definido en (2.24). Entonces, para $t \in \mathbb{R}$, $\widehat{R}^{\tilde{T}_{n_1, n_2}^2}$, satisface que*

$$\sup_{t \in \mathbb{R}} |\widehat{R}^{\tilde{T}_{n_1, n_2}^2}(t) - \chi_p^2(t)| \xrightarrow{p} 0,$$

donde χ_p^2 denota la distribución chi-cuadrado con p grados de libertad.

Dado que el estadístico de contraste T^2 de Hotelling modificado se basa en una norma euclidiana, el test resultante está diseñado para evaluar alternativas de dos colas. Así, concluimos que, para un nivel de significación dado α , se rechaza la hipótesis nula si el estadístico \tilde{T}_{n_1, n_2}^2 se encuentra en el $100\alpha\%$ superior de la distribución de permutaciones.

Capítulo 3

Resultados y discusión

Este capítulo tiene como objetivo implementar las distintas metodologías detalladas en el Capítulo 2. En la Sección 3.1 se describen los conjuntos de datos empleados en las distintas etapas, así como el preprocesado y la preparación de los mismos. En las Secciones 3.2 y 3.3 se analizará si las inyecciones de los ratones son homogéneas mediante dos enfoques complementarios. Finalmente, en la Sección 3.4 se llevará a cabo la comparación del patrón de conectividad a escala de todo el cerebro, partiendo únicamente de aquellos individuos que verifican la condición anterior.

3.1. Preprocesado y preparación de los datos

Base de datos de recuentos celulares

La base de datos de recuentos celulares dispone principalmente de los conteos realizados en las diferentes regiones cerebrales (filas) para cada uno de los ratones (columnas), junto con otras características adicionales, como el volumen o el hemisferio al que pertenecen las regiones. En el Apéndice A se describen en detalle los atributos presentes en la misma, la cual consta de 875 filas, mostrando además un extracto de las cinco primeras filas (véase Tabla A.1), con el objetivo de facilitar su comprensión.

En particular, este conjunto de datos se empleará en la Fase Ia y en la Fase II del experimento, realizándose previamente un proceso de preprocesamiento de los datos, cuyas etapas se detallan a continuación.

1. **Filtrado:** en primer lugar, se eliminarán las mediciones (filas) realizadas sobre la capa I (es decir, aquellas filas cuyo `name` contenga la subpalabra *layer 1*), ya que las cuantificaciones en esta capa no se consideran fiables debido a su baja densidad neuronal y alta diversidad de tipos celulares. Además, se excluirán las mediciones realizadas sobre los tractos y fibras nerviosas (filas cuyo `name` empieza por minúscula), así como aquellas etiquetadas como *universe*, correspondientes a regiones situadas fuera del cerebro que introducen ruido en el análisis. Finalmente, se filtrarán las filas que no verifiquen que todos los recuentos celulares sean mayores que cero, es decir, aquellas que no cumplen que `Controls_counts_i > 0` y `Deprived_counts_j > 0`, para todo $i \in \{81, 82, 95, 96, 100\}$ y $j \in \{104, 105, 107, 114\}$.
2. **Colapso:** una vez filtrados los datos, se procederá a colapsar las regiones que se encuentran desagregadas en las capas II/III, IV, V, VIa y VIb (es decir, aquellas filas cuyo `name` contiene las subpalabras *layer 2/3*, *layer 4*, *layer 5*, *layer 6a* o *layer 6b*),

combinándolas en una única región, excepto en el caso del *barrel field* derecho, que se mantiene sin colapsar. En particular, los recuentos celulares de la región resultante se calculan como la suma del número de neuronas de las diferentes capas para cada uno de los individuos. De este modo, la base de datos final queda constituida por 137 filas, de las cuales 132 corresponden a regiones colapsadas distintas del *barrel field* derecho, mientras que las 5 restantes se asocian a las 5 capas del *barrel field* derecho.

3. **Renombramiento y trasposición:** tras colapsar los datos, se procederá a renombrar las regiones, añadiendo el prefijo RH si pertenecen al hemisferio derecho (`hemisphere=0`) y LH si pertenecen al hemisferio izquierdo (`hemisphere=255`). Seguidamente, se seleccionarán exclusivamente las 9 columnas correspondientes a los recuentos celulares y se realizará la trasposición de la base de datos, de modo que los individuos queden representados por filas y las regiones (variables) por columnas. Finalmente, se creará una nueva variable indicadora del grupo experimental, distinguiéndose entre individuos control (los cinco primeros) e individuos deprivados (los restantes).

De este modo, la base de datos obtenida tras estas etapas, y que se empleará en gran parte del trabajo, queda finalmente representada como una matriz de 9 filas y 138 columnas (incluyendo la variable categórica *grupo experimental*). En concreto, en la Tabla A.2 del Apéndice A se muestra un extracto de la base de datos resultante.

Bases de datos de las coordenadas celulares

Por otro lado, cada individuo dispone de una base de datos de coordenadas celulares, con información sobre las coordenadas espaciales (x, y, z) y las coordenadas transformadas (x_t, y_t, z_t) ¹ de cada una de las neuronas, así como la región anatómica a la que pertenecen. Para facilitar la comprensión de este conjunto de datos, en el Apéndice A se detallan los atributos correspondientes y en la Tabla A.3 se muestra un extracto de las seis primeras filas correspondientes al individuo 81.

Este conjunto de datos se empleará en la Fase Ib del experimento, considerando las coordenadas transformadas de las neuronas para analizar la distribución espacial de los individuos.

3.2. Fase Ia: comparación entre capas de la región *barrel field*

En esta sección, se comprobará si las inyecciones en los ratones son homogéneas mediante un primer enfoque basado en modelizar el número de neuronas (variable respuesta, denotada por Y) según las capas del *barrel field* derecho (variable explicativa categórica) entre los diferentes ratones. Como se mencionó en la Sección 2.1, el estudio se realiza sobre 9 individuos, a partir de los cuales se obtienen 5 mediciones del número de neuronas, correspondientes a las capas corticales $L_{2/3}$, L_4 , L_5 , L_{6a} y L_{6b} , siendo Y_{ij} , el número de

¹Las coordenadas espaciales transformadas de las neuronas son las posiciones anatómicas de dichas neuronas una vez alineadas con el *Allen Brain Atlas*, un espacio de referencia común para todos los cerebros. De este modo, se pueden comparar las posiciones anatómicas de las neuronas entre diferentes cerebros de forma consistente.

neuronas observado en la medición i del ratón j .

Así, para modelizar esta estructura de dependencia, se comenzarán planteando los modelos GLMM con intercepto aleatorio, tanto con distribución de Poisson como Binomial Negativa, junto con su discusión (Sección 3.2.1). Seguidamente, en la Sección 3.2.2 se añadirá una pendiente aleatoria y se volverán a analizar y discutir los modelos, incorporando posteriormente una variable de segundo nivel. Finalmente, en la Sección 3.2.3 se validará el modelo seleccionado y se realizarán las predicciones del número de neuronas para cada ratón según la capa cortical.

3.2.1. Modelos GLMM con intercepto aleatorio para el conteo de neuronas

Modelo de Poisson con intercepto aleatorio

En primer lugar, el modelo de Poisson con intercepto aleatorio viene dado por

$$\log(\lambda_{ij}) = \gamma_{00}^{L_{2/3}} + \alpha_{00}^{L_4} + \alpha_{00}^{L_5} + \alpha_{00}^{L_{6a}} + \alpha_{00}^{L_{6b}} + b_{0j}, \quad j \in \{1, \dots, 9\}, i \in \{1, \dots, n_j\}, \quad (3.1)$$

donde $\gamma_{00}^{L_{2/3}}$, $\alpha_{00}^{L_4}$, $\alpha_{00}^{L_5}$, $\alpha_{00}^{L_{6a}}$ y $\alpha_{00}^{L_{6b}}$ son los efectos fijos asociados a las diferentes capas corticales. Asimismo, la capa $L_{2/3}$ se toma como categoría de referencia, asumiéndose que los efectos aleatorios asociados a ella son independientes y se distribuyen como $b_{0j} \sim N(0, \sigma_{b_0}^2)$.

De lo anterior se deduce que

$$\lambda_{ij} = \exp \left\{ \gamma_{00}^{L_{2/3}} + \alpha_{00}^{L_4} + \alpha_{00}^{L_5} + \alpha_{00}^{L_{6a}} + \alpha_{00}^{L_{6b}} + b_{0j} \right\},$$

por lo que el número esperado de neuronas para la medición realizada en la capa $L_{2/3}$ del individuo j será $\exp\{\gamma_{00}^{L_{2/3}} + b_{0j}\}$, mientras que para la capa k , $k \in \{L_4, L_5, L_{6a}, L_{6b}\}$, será $\exp\{\gamma_{00}^{L_{2/3}} + \alpha_{00}^k + b_{0j}\}$.

En la Tabla 3.1, se muestran las estimaciones de los efectos fijos del modelo (3.1), junto con sus errores típicos y el valor del estadístico de contraste, empleado para comprobar su significación. Vemos que los p -valores obtenidos en los contrastes de significación de los efectos fijos son inferiores a 0.05, por lo que son significativos a un nivel del 5%.

Además, se muestra la estimación de la desviación típica del intercepto (o del efecto aleatorio), $\hat{\sigma}_{b_0} = 0.4087$, valor que se aleja de cero, lo que evidencia la existencia de variabilidad entre individuos en el número esperado de neuronas.

Parámetro	Estimación	Error Estándar	Estadístico z	p-valor
$\gamma_{00}^{L_{2/3}}$	8.7553	0.1363	64.23	$< 2 \cdot 10^{-16}$
$\alpha_{00}^{L_4}$	0.4872	0.0051	95.50	$< 2 \cdot 10^{-16}$
$\alpha_{00}^{L_5}$	-0.8062	0.0072	-111.56	$< 2 \cdot 10^{-16}$
$\alpha_{00}^{L_{6a}}$	-0.7677	0.0071	-107.64	$< 2 \cdot 10^{-16}$
$\alpha_{00}^{L_{6b}}$	-2.3913	0.0139	-172.44	$< 2 \cdot 10^{-16}$
σ_{b_0}	0.4087			

Tabla 3.1. Resumen de la estimación de los parámetros del modelo de Poisson con intercepto aleatorio.

No obstante, para que este modelo resulte adecuado, es necesario verificar la condición de equidispersión. Para ello, se lleva a cabo la estimación del parámetro de sobredispersión del modelo según (2.4), obteniendo un valor de $\hat{\phi} = 872.78$. Dado que $\hat{\phi} > 1$, entonces existe un fenómeno de sobredispersión y, por ende, el modelo no resultará adecuado. Como solución a este problema, se plantea que la variable respuesta siga una distribución Binomial Negativa.

Modelo Binomial Negativo con intercepto aleatorio

En este contexto, se considera el modelo Binomial Negativo con intercepto aleatorio, cuya expresión viene dada por

$$\log(\mu_{ij}) = \gamma_{00}^{L_{2/3}} + \alpha_{00}^{L_4} + \alpha_{00}^{L_5} + \alpha_{00}^{L_{6a}} + \alpha_{00}^{L_{6b}} + b_{0j}, \quad j \in \{1, \dots, 9\}, i \in \{1, \dots, n_j\}, \quad (3.2)$$

con $b_{0j} \sim N(0, \sigma_{b_0}^2)$ independientes, y considerando de nuevo a la capa $L_{2/3}$ como categoría de referencia.

Así, se obtiene que

$$\mu_{ij} = \exp \left\{ \gamma_{00}^{L_{2/3}} + \alpha_{00}^{L_4} + \alpha_{00}^{L_5} + \alpha_{00}^{L_{6a}} + \alpha_{00}^{L_{6b}} + b_{0j} \right\},$$

de modo que, el número esperado de neuronas para una medición realizada en la capa $L_{2/3}$ del individuo j será, como en el caso anterior, $\exp\{\gamma_{00}^{L_{2/3}} + b_{0j}\}$, mientras que, para la capa k , $k \in \{L_4, L_5, L_{6a}, L_{6b}\}$, será $\exp\{\gamma_{00}^{L_{2/3}} + \alpha_{00}^k + b_{0j}\}$.

En la Tabla 3.2, se muestran las estimaciones de los efectos fijos del modelo (3.2), junto con sus errores típicos y el valor del estadístico de contraste. Asimismo, vemos que los p -valores obtenidos en los contrastes de significación de los efectos fijos $\gamma_{00}^{L_{2/3}}$, $\alpha_{00}^{L_5}$, $\alpha_{00}^{L_{6a}}$ y $\alpha_{00}^{L_{6b}}$ son inferiores a 0.05, lo que indica que son significativos a un nivel del 5%, mientras que para el efecto $\alpha_{00}^{L_4}$ no se alcanza dicho umbral de significación, y por tanto, el número de neuronas esperado en la capa L_4 del individuo j , no difiere significativamente

del número de neuronas de la capa de referencia $L_{2/3}$.

Por otro lado, se muestra la estimación de la desviación típica del efecto aleatorio, $\hat{\sigma}_{b_0} = 0.1971$, cuyo valor evidencia la existencia de variabilidad entre individuos en el número esperado de neuronas. Por tanto, la inclusión del efecto aleatorio en el modelo será fundamental para capturar estas diferencias individuales.

Parámetro	Estimación	Error Estándar	Estadístico z	p-valor
$\gamma_{00}^{L_{2/3}}$	8.8001	0.2122	41.475	$< 2 \cdot 10^{-16}$
$\alpha_{00}^{L_4}$	0.4783	0.2826	1.693	0.0905
$\alpha_{00}^{L_5}$	-0.8223	0.2845	-2.890	0.0039
$\alpha_{00}^{L_{6a}}$	-0.7338	0.2825	-2.598	0.0094
$\alpha_{00}^{L_{6b}}$	-2.3743	0.2850	-8.330	$< 2 \cdot 10^{-16}$
σ_{b_0}	0.1971			
θ	2.8296			

Tabla 3.2. Resumen de la estimación de los parámetros del modelo Binomial Negativo con intercepto aleatorio.

Finalmente, comprobemos que el modelo GLMM Binomial Negativo será el más apropiado para nuestros datos. En efecto, consideremos el siguiente contraste:

$$\begin{cases} H_0 : \text{modelo de Poisson con intercepto aleatorio (3.1),} \\ H_a : \text{modelo Binomial Negativo con intercepto aleatorio (3.2).} \end{cases}$$

Para ello, emplearemos el test de razón de verosimilitudes, cuyo estadístico de contraste puede obtenerse a partir del valor de la log-verosimilitud de cada modelo. Tomando la función $\log(L_0)$, para el modelo de Poisson y $\log(L_1)$, para el modelo Binomial, entonces el estadístico de contraste viene dado por

$$T = -2 \log \left(\frac{L_0}{L_1} \right) = -2 \log(L_0) + 2 \log(L_1), \quad (3.3)$$

que bajo la hipótesis nula tiene una distribución χ^2 con m grados de libertad, donde m se obtiene como la diferencia del número de parámetros entre los dos modelos.

En la Tabla 3.3, se muestra para ambos modelos, el número de parámetros, los valores de AIC (Criterio de Información de Akaike) y BIC (Criterio de Información Bayesiano), y el valor de la función de verosimilitud. Además, se proporciona el valor del estadístico de contraste T , obtenido al reemplazar $\log(L_0) = -17685.5$ y $\log(L_1) = -402.2$ en la expresión (3.3), y que vemos que sigue una distribución χ^2 con 1 grado de libertad.

Finalmente, observamos que el p -valor asociado al contraste es inferior a 0.05, por lo que rechazamos la hipótesis nula para un nivel de significación del 5%, lo que indica que seleccionaremos el modelo Binomial Negativo con intercepto aleatorio.

Modelo	N° de parámetros	AIC	BIC	$\log(L_i)$
modelo H_0	6	35 383	35 394	-17 685.5
modelo H_a	7	818	831	-402.2
Estadístico T				34 567
Distribución estadístico T				χ_1^2
p-valor				$< 2 \cdot 10^{-16}$

Tabla 3.3. Comparativa de los modelos GLMM de Poisson (3.1) y Binomial Negativo (3.2).

3.2.2. Modelos GLMM con intercepto y pendiente aleatorios para el conteo de neuronas

Por otro lado, podríamos plantearnos incluir efectos aleatorios distintos para las capas de cada individuo j , tanto en el modelo de Poisson como en el Binomial Negativo con intercepto aleatorio. Esto permitiría capturar, además de la variabilidad entre individuos, las diferencias específicas entre capas.

Modelo de Poisson con intercepto y pendiente aleatorios

En primer lugar, consideremos efectos aleatorios diferentes para las capas de los individuos en el modelo de Poisson con intercepto aleatorio (3.1), obteniendo así la siguiente formulación:

$$\log(\lambda_{ij}) = \gamma_{00}^{L_{2/3}} + \alpha_{00}^{L_4} + \alpha_{00}^{L_5} + \alpha_{00}^{L_{6a}} + \alpha_{00}^{L_{6b}} + b_{0j}^{L_{2/3}} + b_{1j}^{L_4} + b_{1j}^{L_5} + b_{1j}^{L_{6a}} + b_{1j}^{L_{6b}}, \quad (3.4)$$

con $j \in \{1, \dots, 9\}$ e $i \in \{1, \dots, n_j\}$. En particular, los efectos aleatorios asociados a la capa de referencia $L_{2/3}$ son independientes y se distribuyen como $b_{0j}^{L_{2/3}} \sim N(0, \sigma_{b_0}^2)$, y los correspondientes a las restantes capas $k \in \{L_4, L_5, L_{6a}, L_{6b}\}$, son igualmente independientes y se modelan como $b_{1j}^k \sim N(0, \sigma_{b_1^k}^2)$.

De lo anterior se deduce que

$$\lambda_{ij} = \exp \left\{ \gamma_{00}^{L_{2/3}} + \alpha_{00}^{L_4} + \alpha_{00}^{L_5} + \alpha_{00}^{L_{6a}} + \alpha_{00}^{L_{6b}} + b_{0j}^{L_{2/3}} + b_{1j}^{L_4} + b_{1j}^{L_5} + b_{1j}^{L_{6a}} + b_{1j}^{L_{6b}} \right\},$$

por lo que el número esperado de neuronas para la medición realizada en la capa $L_{2/3}$ del individuo j será $\exp\{\gamma_{00}^{L_{2/3}} + b_{0j}^{L_{2/3}}\}$, mientras que para la capa k , con $k \in \{L_4, L_5, L_{6a}, L_{6b}\}$, será $\exp\{\gamma_{00}^{L_{2/3}} + \alpha_{00}^k + b_{0j}^{L_{2/3}} + b_{1j}^k\}$.

En la Tabla 3.4, vemos que salvo para $\alpha_{00}^{L_4}$, los p -valores asociados a los efectos fijos son inferiores a 0.05, y por tanto, significativos a un nivel de significación del 5%. Además, los

valores de las desviaciones típicas de los efectos aleatorios muestran que existe variabilidad entre individuos en las diferentes capas, especialmente en las capas L_5 y L_{6b} . Por ello, podría resultar razonable la inclusión de efectos aleatorios diferenciados entre las capas de los individuos.

Parámetro	Estimación	Error Estándar	Estadístico z	p-valor
$\gamma_{00}^{L_{2/3}}$	8.6163	0.2377	36.251	$< 2 \cdot 10^{-16}$
$\alpha_{00}^{L_4}$	0.5469	0.2854	1.916	0.0553
$\alpha_{00}^{L_5}$	-0.8455	0.3315	-2.550	0.0108
$\alpha_{00}^{L_{6a}}$	-0.6140	0.1795	-3.420	0.0006
$\alpha_{00}^{L_{6b}}$	-2.4943	0.4138	-6.028	$1.66 \cdot 10^{-9}$
σ_{b_0}	0.7129			
$\sigma_{b_1}^{L_4}$	0.8559			
$\sigma_{b_1}^{L_5}$	0.9940			
$\sigma_{b_1}^{L_{6a}}$	0.5381			
$\sigma_{b_1}^{L_{6b}}$	1.2395			

Tabla 3.4. Resumen de la estimación de los efectos fijos y de las desviaciones típicas de los efectos aleatorios del modelo de Poisson con intercepto y pendiente aleatorios.

Por otro lado, el coeficiente de sobredispersión estimado de este modelo es $\hat{\phi} = 0.985$, valor muy cercano a uno, que refleja el cumplimiento de la condición de equidispersión, a diferencia de lo que ocurría para el modelo de Poisson con sólo intercepto aleatorio (3.1). De nuevo, todo parece indicar que habrá que incluir efectos aleatorios distintos entre las capas de los individuos.

Para comprobarlo formalmente, consideremos el siguiente contraste:

$$\begin{cases} H_0 : \text{modelo de Poisson con intercepto aleatorio (3.1)} \\ H_a : \text{modelo de Poisson con intercepto y pendiente aleatorios (3.4),} \end{cases}$$

empleando nuevamente el test de razón de verosimilitudes.

En la Tabla 3.5, vemos que el estadístico de contraste toma el valor $T = 34\,605$ y que sigue una distribución χ^2 con 14 grados de libertad. Con un p -valor $< 2 \cdot 10^{-16}$, rechazamos la hipótesis nula para un nivel de significación del 5%, por lo que nos quedaremos con el modelo de Poisson con intercepto y pendiente aleatorios.

Modelo	N° de parámetros	AIC	BIC	log(L _i)
modelo H_0	6	35 383	35 394	-17 686
modelo H_a	20	806	842	-383
Estadístico T				34 605
Distribución estadístico T				χ_{14}^2
p-valor				$< 2 \cdot 10^{-16}$

Tabla 3.5. Comparativa de los modelos GLMM de Poisson (3.1) y (3.4).

Modelo Binomial Negativo con intercepto y pendiente aleatorios

Ahora, consideraremos efectos aleatorios diferentes para las capas de los individuos en el modelo Binomial Negativo con intercepto aleatorio (3.2). Así, se obtiene la siguiente formulación:

$$\log(\mu_{ij}) = \gamma_{00}^{L_{2/3}} + \alpha_{00}^{L_4} + \alpha_{00}^{L_5} + \alpha_{00}^{L_{6a}} + \alpha_{00}^{L_{6b}} + b_{0j}^{L_{2/3}} + b_{1j}^{L_4} + b_{1j}^{L_5} + b_{1j}^{L_{6a}} + b_{1j}^{L_{6b}}, \quad (3.5)$$

con $j \in \{1, \dots, 9\}$, $i \in \{1, \dots, n_j\}$, $b_{0j}^{L_{2/3}} \sim N(0, \sigma_{b_0}^2)$ independientes y para cada capa $k \in \{L_4, L_5, L_{6a}, L_{6b}\}$, $b_{1j}^k \sim N(0, \sigma_{b_1^k}^2)$ independientes.

De este modo, se deduce que

$$\mu_{ij} = \exp \left\{ \gamma_{00}^{L_{2/3}} + \alpha_{00}^{L_4} + \alpha_{00}^{L_5} + \alpha_{00}^{L_{6a}} + \alpha_{00}^{L_{6b}} + b_{0j}^{L_{2/3}} + b_{1j}^{L_4} + b_{1j}^{L_5} + b_{1j}^{L_{6a}} + b_{1j}^{L_{6b}} \right\},$$

donde el número esperado de neuronas para las mediciones realizadas en las distintas capas del ratón j se calcula de forma análoga al caso anterior.

En la Tabla 3.6, vemos que los p -valores asociados a todos los efectos fijos son inferiores a 0.05, y por tanto, significativos a un nivel de significación del 5%. De nuevo, los valores de las desviaciones típicas de los efectos aleatorios muestran que existe variabilidad entre individuos en las diferentes capas, lo que indica que podría resultar razonable la inclusión de efectos aleatorios diferenciados entre las capas de los individuos.

Parámetro	Estimación	Error Estándar	Estadístico z	p-valor
$\gamma_{00}^{L_{2/3}}$	8.6161	0.0861	99.975	$< 2 \cdot 10^{-16}$
$\alpha_{00}^{L_4}$	0.5469	0.0853	6.412	$1.44 \cdot 10^{-10}$
$\alpha_{00}^{L_5}$	-0.8453	0.1069	-7.904	$2.69 \cdot 10^{-15}$
$\alpha_{00}^{L_{6a}}$	-0.6138	0.0802	-7.655	$1.93 \cdot 10^{-14}$
$\alpha_{00}^{L_{6b}}$	-2.4942	0.2893	-8.622	$< 2 \cdot 10^{-16}$
σ_{b_0}	0.7129			
$\sigma_{b_1}^{L_4}$	0.8559			
$\sigma_{b_1}^{L_5}$	0.9940			
$\sigma_{b_1}^{L_{6a}}$	0.5381			
$\sigma_{b_1}^{L_{6b}}$	1.2395			
θ	26 700 794			

Tabla 3.6. Resumen de la estimación de los efectos fijos, de las desviaciones típicas de los efectos aleatorios y del parámetro θ del modelo Binomial Negativo con intercepto y pendiente aleatorios.

Veamos, que en efecto, habrá que incluir efectos aleatorios distintos entre las capas de los individuos. Para ello, consideremos el siguiente contraste:

$$\begin{cases} H_0 : \text{modelo Binomial Negativo con intercepto aleatorio (3.2),} \\ H_a : \text{modelo Binomial Negativo con intercepto y pendiente aleatorios (3.5).} \end{cases}$$

Empleando el test de razón de verosimilitudes, vemos en el Cuadro 3.7, que el estadístico de contraste toma el valor $T = 38.328$ y que sigue una distribución χ^2 con 14 grados de libertad. Con un p -valor 0.00046, rechazamos la hipótesis nula para un nivel de significación del 5%, por lo que nos quedaremos con el modelo Binomial Negativo con intercepto y pendiente aleatorios.

Modelo	N° de parámetros	AIC	BIC	$\log(\mathbf{L}_i)$
modelo H_0	7	818.38	831.02	-402.19
modelo H_a	21	808.05	845.99	-383.02
Estadístico T				38.328
Distribución estadístico T				χ_{14}^2
p-valor				0.00046

Tabla 3.7. Comparativa de los modelos GLMM Binomial Negativo (3.2) y (3.5).

Por otro lado, en la Tabla 3.6 vemos que el parámetro de sobredispersión θ es muy elevado, de modo que la varianza condicional de la variable respuesta

$$\text{Var}(Y_{ij} | \mathbf{b}_j) = \mu_{ij} + \frac{\mu_{ij}^2}{\theta},$$

se aproxima a la de un modelo de Poisson (la cual es exactamente igual a la media condicional $\mathbb{E}(Y_{ij} | \mathbf{b}_j) = \mu_{ij}$). Esto refleja que podría ser razonable considerar el modelo de Poisson en vez de el Binomial Negativo con intercepto y pendiente aleatorios.

Para verificarlo, consideremos el siguiente contraste:

$$\begin{cases} H_0 : \text{modelo de Poisson con intercepto y pendiente aleatorios (3.4),} \\ H_a : \text{modelo Binomial Negativo con intercepto y pendiente aleatorios (3.5).} \end{cases}$$

Empleando de nuevo el test de razón de verosimilitudes, vemos a continuación, que con un p -valor de 1, no tenemos evidencias significativas en contra de la hipótesis nula, por lo que nos quedaremos con el modelo de Poisson con intercepto y pendiente aleatorios (3.4).

Modelo	N° de parámetros	AIC	BIC	$\log(\mathbf{L}_i)$
modelo H_0	20	806.05	842.18	-383.02
modelo H_a	21	808.05	845.99	-383.02
Estadístico T				0
Distribución estadístico T				χ_1^2
p-valor				1

Tabla 3.8. Comparativa de los modelos de Poisson y Binomial Negativo con intercepto y pendiente aleatorios.

Inclusión de una variable de segundo nivel en el modelo de Poisson con intercepto y pendientes aleatorios

Una vez considerado el modelo de Poisson con intercepto y pendientes aleatorios (3.4), nos podríamos plantear introducir la variable explicativa de segundo nivel $W = \text{Grupo}$ (distinguiendo entre control y deprivado). De este modo, consideramos la siguiente formulación:

$$\log(\lambda_{ij}) = \gamma_{00}^{L_{2/3}} + \alpha_{00}^{L_4} + \alpha_{00}^{L_5} + \alpha_{00}^{L_{6a}} + \alpha_{00}^{L_{6b}} + \delta_{00}^D + b_{0j}^{L_{2/3}} + b_{1j}^{L_4} + b_{1j}^{L_5} + b_{1j}^{L_{6a}} + b_{1j}^{L_{6b}}, \quad (3.6)$$

con $j \in \{1, \dots, 9\}$, $i \in \{1, \dots, n_j\}$, $b_{0j}^{L_{2/3}} \sim N(0, \sigma_{b_0}^2)$ independientes y para cada capa $k \in \{L_4, L_5, L_{6a}, L_{6b}\}$, $b_{1j}^k \sim N(0, \sigma_{b_1^k}^2)$ independientes.

Así, obtenemos que

$$\lambda_{ij} = \exp \left\{ \gamma_{00}^{L_{2/3}} + \alpha_{00}^{L_4} + \alpha_{00}^{L_5} + \alpha_{00}^{L_{6a}} + \alpha_{00}^{L_{6b}} + \delta_{00}^D + b_{0j}^{L_{2/3}} + b_{1j}^{L_4} + b_{1j}^{L_5} + b_{1j}^{L_{6a}} + b_{1j}^{L_{6b}} \right\},$$

de modo que en la Tabla 3.9 se muestra el número de neuronas para las distintas capas de los individuos, distinguiendo según sean control ó deprivados.

Escenario	N° esperado de neuronas
Capa $L_{2/3}$ de indiv. control	$\exp\{\gamma_{00}^{L_{2/3}} + b_{0j}^{L_{2/3}}\}$
Capa $L_{2/3}$ de indiv. deprivado	$\exp\{\gamma_{00}^{L_{2/3}} + \delta_{00}^D + b_{0j}^{L_{2/3}}\}$
Capa k , $k \in \{L_4, L_5, L_{6a}, L_{6b}\}$ de indiv. control	$\exp\{\gamma_{00}^{L_{2/3}} + \alpha_{00}^k + b_{0j}^{L_{2/3}} + b_{1j}^k\}$
Capa k , $k \in \{L_4, L_5, L_{6a}, L_{6b}\}$ de indiv. deprivado	$\exp\{\gamma_{00}^{L_{2/3}} + \alpha_{00}^k + \delta_{00}^D + b_{0j}^{L_{2/3}} + b_{1j}^k\}$

Tabla 3.9. Número esperado de neuronas para los diferentes escenarios.

Las estimaciones de los efectos fijos del modelo (3.6), se muestran en la Tabla 3.10, junto con los p -valores de sus respectivos contrastes de significación. Vemos que los efectos fijos de las capas serán significativos a un nivel de significación del 5%, mientras que para el efecto fijo del grupo, δ_{00}^D , no se alcanza dicho umbral. Por tanto, dada una determinada capa, el número esperado de neuronas en el individuo deprivado, no difiere significativamente del individuo control. Esto sugiere que no sería necesario incluir la variable de segundo nivel en el modelo.

Parámetro	Estimación	Error Estándar	Estadístico z	p-valor
$\gamma_{00}^{L_{2/3}}$	8.6309	0.24311	35.501	$< 2 \cdot 10^{-16}$
$\alpha_{00}^{L_4}$	0.5468	0.28538	1.916	0.0454
$\alpha_{00}^{L_5}$	-0.8456	0.33151	-2.551	0.0107
$\alpha_{00}^{L_{6a}}$	-0.6140	0.17954	-3.420	0.00063
$\alpha_{00}^{L_{6b}}$	-2.4943	0.41379	-6.028	$1.66 \cdot 10^{-9}$
δ_{00}^D	-0.0326	0.09840	-0.332	0.7401

Tabla 3.10. Resumen de la estimación de los efectos fijos del modelo GLMM de Poisson con intercepto y pendiente aleatorios y variable explicativa de segundo nivel W .

Finalmente, comprobemos tal y como se esperaba, que no será necesario incluir el efecto de la variable de segundo nivel W en el modelo (3.4). En efecto, consideremos el siguiente contraste:

$$\begin{cases} H_0 : \text{modelo de Poisson sin la variable de segundo nivel (3.4),} \\ H_a : \text{modelo de Poisson con la variable de segundo nivel (3.6),} \end{cases}$$

En la Tabla 3.11, vemos que el estadístico de contraste toma el valor $T = 0.1094$ y que sigue una distribución χ^2 con 1 grado de libertad. Con un p -valor de 0.7408, no tenemos evidencias significativas en contra de la hipótesis nula, por lo que nos quedaremos con el modelo de Poisson con intercepto y pendiente aleatoria sin incluir el efecto de la variable de segundo nivel W (3.4).

Modelo	N° de parámetros	AIC	BIC	$\log(L_i)$
modelo H_0	20	806.05	842.18	-383.02
modelo H_a	21	807.94	845.88	-382.97
Estadístico T				0.1094
Distribución estadístico T				χ_1^2
p-valor				0.7408

Tabla 3.11. Comparativa de los modelos GLMM de Poisson (3.4) y (3.6).

3.2.3. Validación y predicciones del modelo de Poisson con intercepto y pendiente aleatorios

Una vez seleccionado el modelo de Poisson con intercepto y pendiente aleatorios (3.4), se procede a su correspondiente validación, es decir, a la verificación del cumplimiento de las hipótesis. Para ello, será necesario comprobar si los efectos aleatorios tienen distribución normal. Es decir, dado $j \in \{1, \dots, 9\}$, probemos que $b_{0j}^{L_{2/3}} \sim N(0, \sigma_{b_0}^2)$, $b_{1j}^{L_4} \sim N(0, \sigma_{b_1^{L_4}}^2)$, $b_{1j}^{L_5} \sim N(0, \sigma_{b_1^{L_5}}^2)$, $b_{1j}^{L_{6a}} \sim N(0, \sigma_{b_1^{L_{6a}}}^2)$ y $b_{1j}^{L_{6b}} \sim N(0, \sigma_{b_1^{L_{6b}}}^2)$.

Luego, para cada capa l , $l \in \{L_{2/3}, L_4, L_5, L_{6a}, L_{6b}\}$, consideremos los siguientes contrastes:

$$\begin{cases} H_0 : \text{los efectos aleatorios de la capa } l \text{ tienen distribución Normal,} \\ H_a : \text{los efectos aleatorios de la capa } l \text{ no tienen distribución Normal,} \end{cases} \quad (3.7)$$

empleando para ello el test de Shapiro-Wilk. Para más información se puede consultar [Salkind, 2007](#).

En la Tabla 3.12, vemos que para los efectos aleatorios de todas las capas, los p -valores son superiores a 0.05, por lo que no tenemos evidencias significativas en contra de la hipótesis nula, y por ende, los efectos tienen distribución normal.

Efectos aleatorios	Estadístico W	p-valor
Capa $L_{2/3}$	0.8089	0.0502
Capa L_4	0.9853	0.9858
Capa L_5	0.9114	0.3255
Capa L_{6a}	0.9681	0.8781
Capa L_{6b}	0.8602	0.0964

Tabla 3.12. Resumen de los contrastes de normalidad para los efectos aleatorios de cada capa.

En la Figura 3.1, se muestran los gráficos Q-Q de normalidad para los efectos aleatorios de cada capa, los cuales comparan los cuantiles observados de los datos con los cuantiles teóricos de una distribución normal. Vemos, como era previsible, que para todas las capas los puntos se sitúan próximos a la recta diagonal de referencia, lo que respalda la hipótesis de normalidad de los efectos aleatorios de las capas corticales.

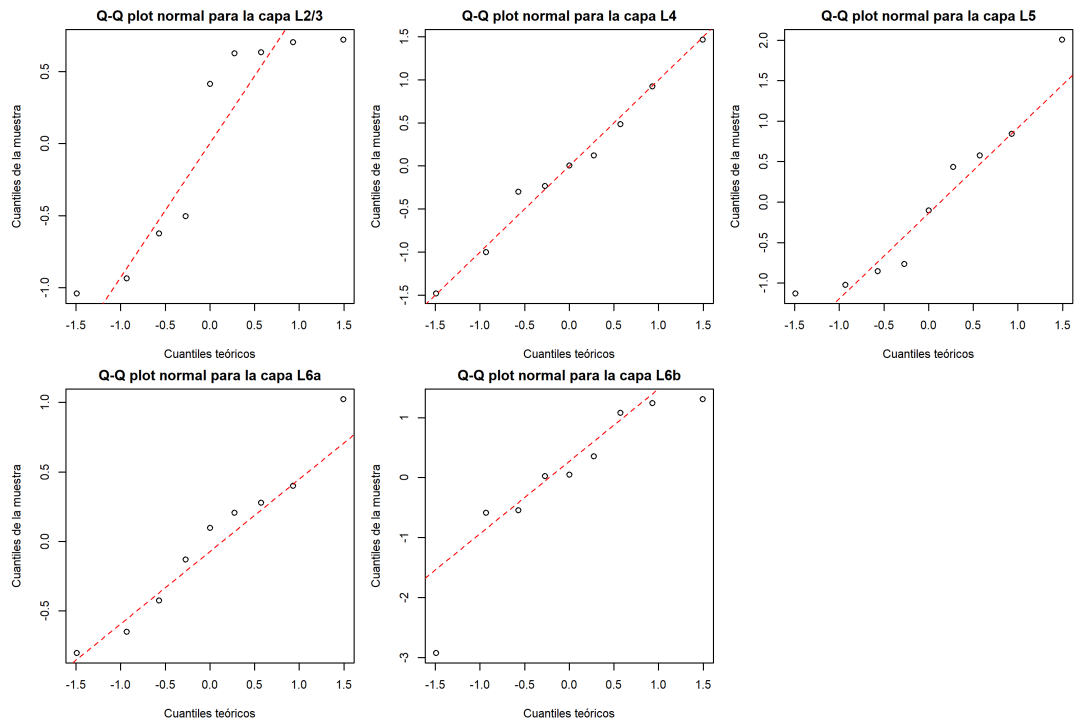


Figura 3.1. Gráficos Q-Q de normalidad para los efectos aleatorios de cada capa.

Una vez validado el modelo (3.4), se procede a ajustar el número de neuronas en las diferentes capas corticales de los individuos. Seguidamente, en la Tabla 3.13 se muestran las estimaciones obtenidas.

	$L_{2/3}$	L_4	L_5	L_{6a}	L_{6b}
Individuo 81	2962.82	8338.29	2262.53	1971.60	723.25
Individuo 82	8353.74	5317.46	1532.82	3965.83	400.57
Individuo 95	10 406.12	18 110.06	6894.99	3686.46	1225.17
Individuo 96	1954.48	14 648.99	6248.99	2945.56	165.694
Individuo 100	3337.07	4573.50	517.04	2385.05	953.92
Individuo 104	11 353.91	22 191.72	4402.82	3209.55	985.68
Individuo 105	10 333.65	4065.52	1437.77	6161.13	45.94
Individuo 107	2170.80	9444.03	2165.89	1753.52	663.85
Individuo 114	11 154.46	14 273.46	2235.13	2707.11	511.85

Tabla 3.13. Estimaciones del número de neuronas por capa cortical para los diferentes individuos, según el modelo de Poisson con intercepto y pendiente aleatorias (3.4).

En la Figura 3.2 se representan gráficamente los ajustes del número de neuronas para cada ratón en función de la capa cortical. Se observa que, salvo en los individuos 82 y 105, existe un patrón muy similar entre los sujetos. En particular, el número esperado de neuronas aumenta desde la capa $L_{2/3}$ hasta la capa L_4 , y posteriormente disminuye de la capa L_4 a la capa L_5 . En cuanto a las capas L_{6a} y L_{6b} , se aprecian ligeras diferencias en el patrón, aunque estas podrían atribuirse a su naturaleza como subcapas de la capa L_6 , por lo que su relevancia interpretativa es limitada. En consecuencia, se concluye que la distribución laminar de las neuronas es comparable entre los ratones, a excepción de los individuos 82 y 105, que presentan un comportamiento notablemente distinto.

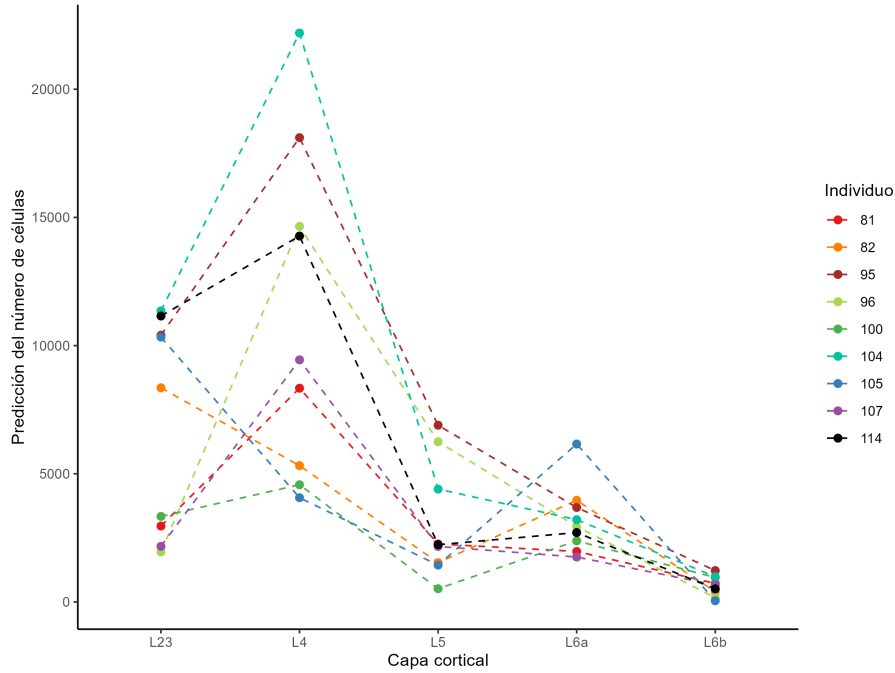


Figura 3.2. Predicciones del número de neuronas por capa cortical para los diferentes individuos, según el modelo de Poisson con intercepto y pendiente aleatorias (3.4).

3.3. Fase Ib: comparación de la distribución espacial celular

En esta sección, se validará si las inyecciones realizadas en los ratones son comparables mediante un segundo enfoque basado en la distribución espacial de las neuronas del *barrel field* derecho.

En primer lugar, para cada ratón $k \in \{1, \dots, 9\}$, denotemos por

$$\mathbf{X}_k = \{(x_{k,1}, y_{k,1}, z_{k,1}), \dots, (x_{k,N_k}, y_{k,N_k}, z_{k,N_k})\},$$

al conjunto de coordenadas celulares transformadas distribuidas dentro de la región de observación tridimensional

$$W = [0, 200] \times [220, 300] \times [40, 140] \subset \mathbb{R}^3,$$

correspondiente al *barrel field* derecho, y donde N_k denota el número total de neuronas registradas sobre el individuo k . En particular, en la Tabla 3.14 se muestra el número de células para cada ratón.

Individuo	Número de neuronas en el <i>barrel field</i>
81	16 415
82	13 533
95	41 347
96	26 117
100	12 595
104	44 424
105	26 451
107	16 611
114	19 848

Tabla 3.14. Número total de neuronas observadas en la región *barrel field* del hemisferio derecho para los diferentes individuos.

En la Figura 3.3 se representa la estimación de la intensidad para cada uno de los ratones analizados, es decir, la cantidad esperada de neuronas por unidad de volumen.

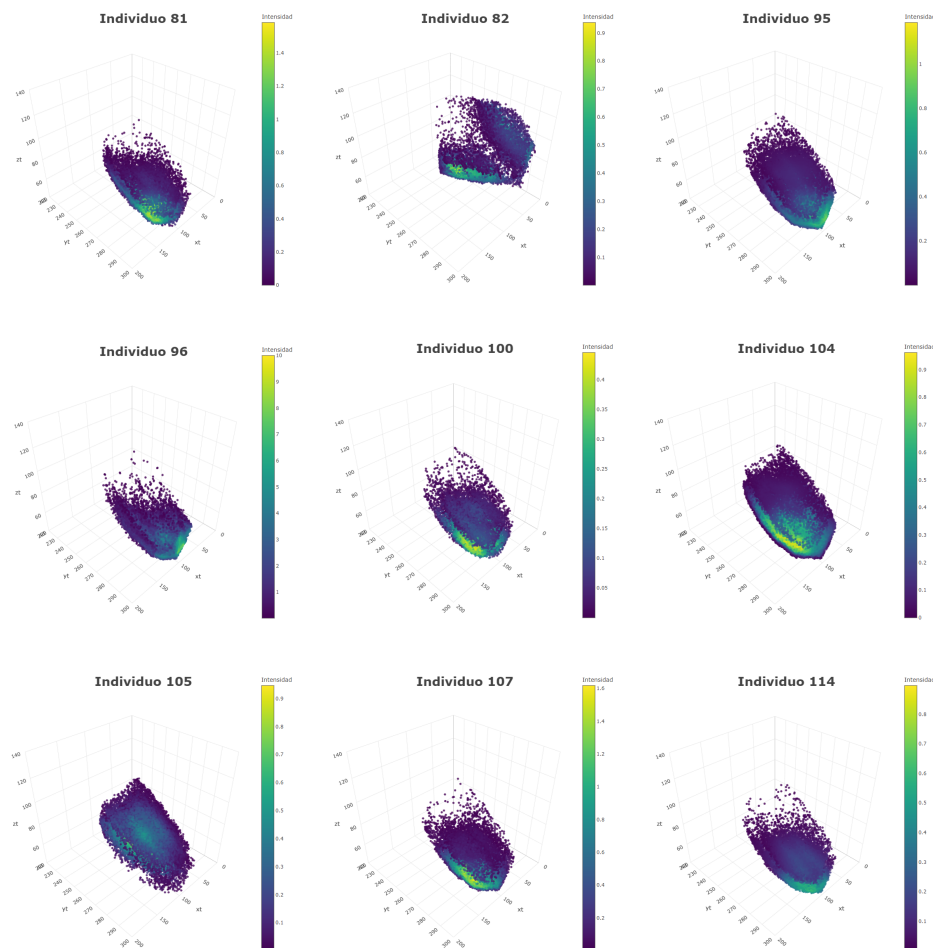


Figura 3.3. Representación tridimensional de la estimación de la intensidad para cada uno de los individuos en la región *barrel field* del hemisferio derecho.

En vista de la Figura 3.3, se observa que existen zonas con mayor concentración de neuronas que otras, lo que implica que la función de intensidad no es constante. Además, para los eventos de cada individuo se asume independencia espacial, pues el marcaje de una neurona no condiciona el marcaje de otras neuronas, independientemente de su proximidad espacial; es decir, la aparición de una neurona marcada no altera la probabilidad de observar otras neuronas en el resto del espacio. Por ello, cada conjunto de coordenadas X_k , puede considerarse una realización de un proceso puntual de Poisson inhomogéneo.

Así, para cada par de individuos $k, l \in \{1, \dots, 9\}$, con $k \neq l$, $\mathbf{X}_k^{(k,l)} = \mathbf{X}_k \cup \mathbf{X}_l$ constituye una realización de un proceso puntual de Poisson bivariado inhomogéneo. A partir de esta construcción, se compararán dos a dos las distribuciones espaciales de los conjuntos de coordenadas asociados a los nueve individuos mediante el contraste de la hipótesis nula

$$H_0 : \lambda_{0k}(x, y, z) = \lambda_{0l}(x, y, z) \text{ para casi todo } (x, y, z) \in W, \quad (3.8)$$

siendo λ_{0k} y λ_{0l} las densidades de localización de eventos de los conjuntos X_k y X_l respectivamente.

Para ello, se recurrirá en primer lugar a la calibración asintótica del test que, de acuerdo con el Teorema 2.18, permite aproximar la distribución normal asintótica del estadístico de contraste \widehat{T} descrito en (3.3). En particular, se empleará la función núcleo gaussiana trivariante y se estimarán las matrices ancho de banda mediante el método *plug-in* para obtener dicha calibración.

Asimismo, dado que se llevan a cabo $\binom{9}{2} = 36$ contrastes dos a dos entre las densidades de localización de los conjuntos de coordenadas, se recurre a la corrección de Bonferroni para evitar la inflación del error de tipo I² asociada a comparaciones múltiples. De este modo, dado el nivel de significación original $\alpha = 0.05$, la corrección de Bonferroni establece que cada hipótesis individual debe contrastarse utilizando un nivel de significación reducido de $\alpha/36 = 0.0014$. En la Tabla 3.15 se muestran los p -valores obtenidos para cada uno de los 36 contrastes mediante la calibración asintótica del test.

Contraste asintótico								
$\alpha_{\text{Bonf}} = 0.0014$	81	82	95	96	100	104	105	107
82	0.0132							
95	0.4177	0.0091						
96	0.2050	0.0012	0.2831					
100	0.4348	0.0004	0.4331	0.2372				
104	0.3835	0.0037	0.3344	0.1029	0.4263			
105	0.1514	0.0513	0.0835	0.0009	0.2173	0.2639		
107	0.4655	0.0059	0.4219	0.2397	0.4101	0.3445	0.0933	
114	0.2867	0.0022	0.2679	0.1610	0.3281	0.1771	0.0170	0.3025

Tabla 3.15. Resumen de los p -valores correspondientes a los 36 contrastes obtenidos a partir de la calibración asintótica del test, representando en azul los inferiores a α_{Bonf} .

²El error de tipo I, también denominado falso positivo, es el error que se comete cuando se rechaza la hipótesis nula H_0 , siendo esta verdadera en la población.

De este modo, observamos que para los contrastes por pares entre los individuos 81, 95, 96, 100, 104, 107 y 114, los p -valores asociados son superiores al nivel de significación establecido $\alpha_{\text{Bonf}} = 0.0014$, por lo que no tenemos evidencias significativas en contra de la hipótesis nula. En consecuencia, se asume que las densidades de localización de eventos entre dichos individuos son iguales dos a dos.

No obstante, para los contrastes por pares entre el individuo 82 y los individuos 96 y 100, los p -valores asociados son inferiores al nivel de significación establecido, de modo que se rechaza la hipótesis nula de igualdad en las densidades de localización de eventos de ambos individuos. Lo mismo sucede para el contraste entre los individuos 105 y 96.

Por otro lado, se emplea la calibración *bootstrap* del test para contrastar la hipótesis nula (3.8), aplicando el método *bootstrap* suavizado para obtener la distribución del estadístico de contraste \hat{T} . Al igual que en el caso anterior, por disponer de 36 contrastes, se establece nuevamente el nivel de significación corregido por Bonferroni $\alpha_{\text{Bonf}} = 0.0014$,

En la Tabla 3.16 se recogen los p -valores obtenidos para cada uno de los 36 contrastes mediante la calibración *bootstrap* del test.

$\alpha_{\text{Bonf}} = 0.0014$	Contraste bootstrap							
	81	82	95	96	100	104	105	107
82		0.0012						
95	0.4680	0.0009						
96	0.4553	0.0007	0.4418					
100	0.4306	0.0014	0.4239	0.3125				
104	0.5127	0.0215	0.4468	0.3561	0.4019			
105	0.0189	0.0674	0.0038	0.0010	0.0069	0.0032		
107	0.5362	0.0008	0.3745	0.4027	0.3896	0.4581	0.0017	
114	0.2873	0.0086	0.1964	0.1428	0.3389	0.1715	0.0011	0.2196

Tabla 3.16. Resumen de los p -valores correspondientes a los 36 contrastes obtenidos a partir de la calibración *bootstrap* del test, representando en color azul los inferiores a α_{Bonf} .

En este caso, observamos de nuevo que para los individuos 81, 95, 96, 100, 104, 107 y 114, las densidades de localización de eventos entre dichos individuos son iguales dos a dos, pues los p -valores asociados son superiores a $\alpha_{\text{Bonf}} = 0.0014$. No obstante, para los contrastes por pares entre el individuo 82 y los individuos 81, 95, 96 y 107, así como entre el individuo 105 y los individuos 96 y 114 no se verifica la hipótesis nula.

En conclusión, tanto la calibración asintótica del test como la aproximación mediante *bootstrap* muestran que, salvo en los individuos 82 y 105, las densidades de localización de eventos no difieren dos a dos entre los demás ratones. Este resultado es consistente con lo obtenido en la Fase Ia, donde la distribución laminar de las neuronas era comparable entre los mismos ratones. En consecuencia, ambos análisis confirman la homogeneidad de las inyecciones, salvo en los individuos 82 y 105, que serán eliminados en la siguiente etapa del experimento.

3.4. Fase II: comparación multivariante celular de los grupos experimentales

3.4.1. Aplicación del análisis de componentes principales

Aunque inicialmente se disponía de 9 ratones, tras la fase de validación, la muestra final queda constituida por 7 ratones; en particular, los individuos del grupo control etiquetados como 81, 95, 96 y 100 y los individuos del grupo deprivado etiquetados como 104, 107 y 114. Para cada uno de ellos se dispone del número de células observado en las 133 regiones cerebrales objeto de estudio, incluyendo la zona de inyección, denotada como `RH_Primary somatosensory area, barrel field`. Dichas regiones se agrupan en las siguientes áreas funcionales, tal y como se muestra en la Tabla B.1. En concreto, se distinguen las áreas visuales, motoras, auditivas, asociativas³, somatosensoriales⁴, olfativas y gustativas, situadas en la corteza cerebral, y por otro lado, las regiones de localización más profunda, que comprenden las áreas límbicas⁵, subcorticales⁶, hipocampales⁷, troncales⁸ y cerebelosas.

Dado que el comportamiento de la inyección en la zona cero condiciona los recuentos celulares del resto de regiones, se opta por trabajar con datos relativos al *barrel field* del hemisferio derecho, en lugar de utilizar los recuentos absolutos. Así, a partir de los datos originales, se considera el vector aleatorio 132-dimensional $\mathbf{X} = (X_1, \dots, X_{132})^\top$, con vector de medias $\boldsymbol{\mu}$ y matriz de covarianzas $\boldsymbol{\Sigma}$, donde cada variable X_i representa el número de células en la región cerebral i normalizado respecto al número de células de la región de inyección. Debido a la alta dimensionalidad del conjunto de datos, surge la necesidad de aplicar el PCA. Sin embargo, para ello, es necesario comprobar la existencia de correlaciones lineales entre las variables y su adecuada comparabilidad en términos de escala.

En particular, la evidencia de correlación entre variables se analiza de forma detallada en la Figura 3.4, donde se representa gráficamente la matriz de correlaciones R . Esta matriz se obtiene calculando, para cada par de variables, el coeficiente de correlación de Pearson r_{ij} , que mide la relación lineal entre ambas y se define como el cociente entre su covarianza y el producto de sus desviaciones típicas. En ella se observan mayoritariamente celdas con tonalidades rojas y azules intermedias, lo que refleja, respectivamente, fuertes correlaciones positivas (mayor que 0.8) y correlaciones negativas de intensidad moderada (entre -0.8 y -0.5). Por el contrario, apenas se aprecian tonalidades claras cercanas al blanco, lo que evidencia la escasez de correlaciones débiles. En conjunto, esta estructura de elevada dependencia entre variables justifica la aplicación del PCA.

³Las áreas asociativas son las regiones corticales encargadas de integrar y procesar información procedente de distintas áreas sensoriales y motoras.

⁴Las áreas somatosensoriales son las regiones corticales encargadas de procesar la información sensitiva procedente del cuerpo, como el tacto, la presión, la temperatura o el dolor.

⁵Las áreas límbicas son las regiones cerebrales implicadas principalmente en la regulación de las emociones y la memoria.

⁶Las áreas subcorticales son las regiones localizadas por debajo de la corteza cerebral que participan en funciones esenciales como el movimiento, la memoria, la regulación emocional y el procesamiento de la información sensorial.

⁷Las áreas hipocampales son las regiones del lóbulo temporal medial implicadas en la formación y consolidación de la memoria.

⁸Las áreas troncales son las regiones localizadas en el tronco encefálico que se encargan de controlar de forma automática funciones básicas e involuntarias, como la respiración, el ritmo cardíaco o la presión arterial.

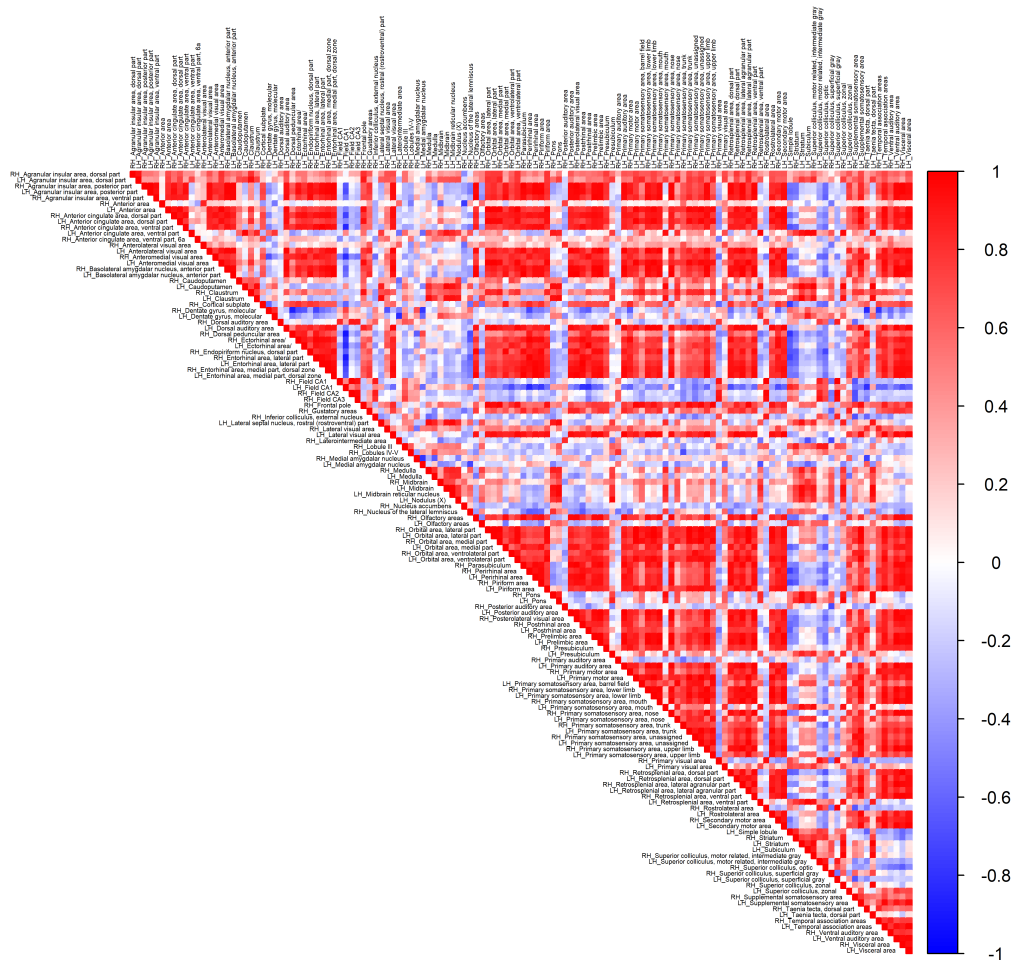


Figura 3.4. Matriz de correlación de Pearson entre las variables asociadas a las distintas regiones cerebrales, definidas como el número de células normalizado respecto a la zona de inyección.

Por otro lado, a pesar de que las variables están definidas en las mismas unidades, presentan diferencias en su magnitud. Por ello, se procede a su estandarización, lo que equivale a trabajar con la matriz de correlaciones de Pearson R , en lugar de la matriz de covarianzas.

Como vimos en la Sección 2.3.1, el objetivo de este apartado será transformar el conjunto de variables aleatorias $\mathbf{X} = (X_1, \dots, X_{132})^\top$ en un nuevo conjunto más pequeño de variables incorreladas que expliquen gran parte de la información de los datos originales.

A continuación, se muestra para cada una de las siete primeras componentes principales, el autovalor correspondiente, es decir, la varianza asociada:

Componente principal	Varianza
Y_1	66.9006
Y_2	29.2742
Y_3	13.9737
Y_4	9.7864
Y_5	7.1080
Y_6	4.9571
Y_7	$8.3718 \cdot 10^{-15}$

Tabla 3.17. Varianza de las siete primeras componentes principales.

De este modo, obtendremos para las siete primeras componentes principales la proporción de variabilidad explicada según (2.17) y la proporción de variabilidad explicada acumulada según (2.18):

Componente principal	Proporción de varianza explicada	Proporción de varianza acumulada
Y_1	0.5068	0.5068
Y_2	0.2218	0.7286
Y_3	0.1059	0.8345
Y_4	0.0741	0.9086
Y_5	0.0538	0.9624
Y_6	0.0375	≈ 1.0000
Y_7	$6.3423 \cdot 10^{-17}$	≈ 1.0000

Tabla 3.18. Proporción de varianza explicada y de varianza explicada acumulada de las siete primeras componentes principales.

En la Tabla 3.18 se observa que la primera componente principal explica por sí sola más del 50 % de la variabilidad total, de modo que captura la información más relevante del conjunto de variables originales. Asimismo, la componente Y_2 contribuye notablemente al incremento de la varianza explicada acumulada. No obstante, a partir de Y_3 , la contribución de cada componente se vuelve muy pequeña, siendo para Y_7 prácticamente nula.

Con el fin de determinar el número adecuado de componentes principales para explicar la variabilidad de los datos, se aplicarán los tres criterios descritos en la Sección 2.3.1, seleccionándose finalmente el más apropiado conforme a los objetivos del estudio.

En primer lugar, por el criterio de la varianza explicada se retienen cuatro componentes principales para explicar el 90 % de la variabilidad total, siendo este el umbral establecido

habitualmente.

Por otro lado, dado que el PCA se realiza sobre la matriz de correlaciones, el criterio de Kaiser establece que deben conservarse únicamente las componentes principales con autovalores superiores a 1, reteniéndose así las seis primeras componentes, tal y como se aprecia en la Tabla 3.17.

Finalmente, en la Figura 3.5 se muestra el gráfico de sedimentación, observándose un “codo” en la componente Y_3 , a partir del cual la curva se estabiliza progresivamente, lo que justifica la retención de las tres primeras componentes principales.

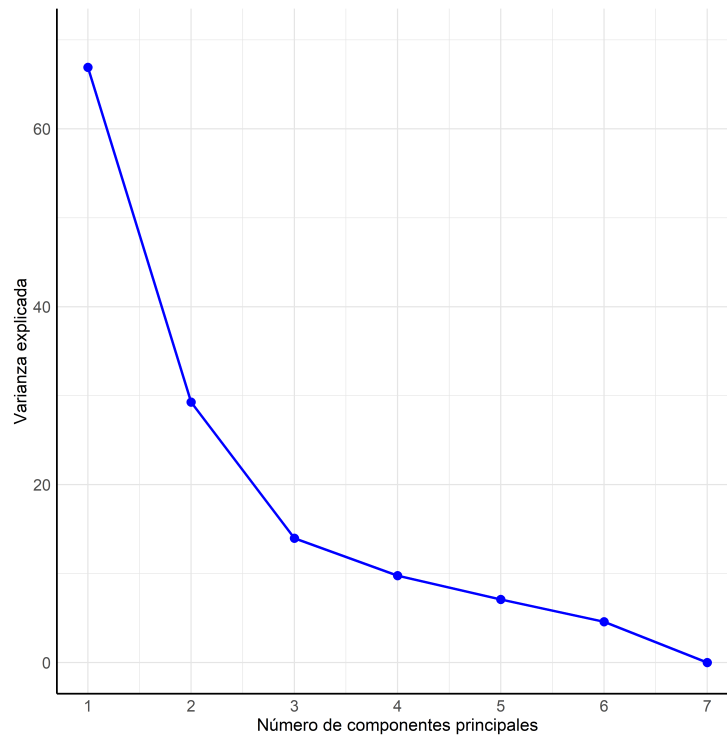


Figura 3.5. Gráfico de sedimentación para las siete primeras componentes principales.

En base a los tres criterios considerados, se observa cierta discrepancia en la selección del número óptimo de componentes principales. No obstante, dado que el análisis se realiza con una muestra muy reducida (únicamente siete individuos), frente a un elevado número de variables, se opta por retener las tres primeras componentes principales, ya que concentran la mayor parte de la variabilidad relevante de los datos y ofrecen una representación más estable.

Por otro lado, en la Figura 3.6 se muestra el gráfico biplot, es decir, la representación conjunta de los individuos (puntos) y de las variables originales (vectores) en el espacio definido por las dos primeras componentes principales, que recordemos que explican la mayor parte de la variabilidad total del conjunto de datos original. Además, las variables se distinguen según el área funcional al que pertenecen mediante una leyenda de colores.

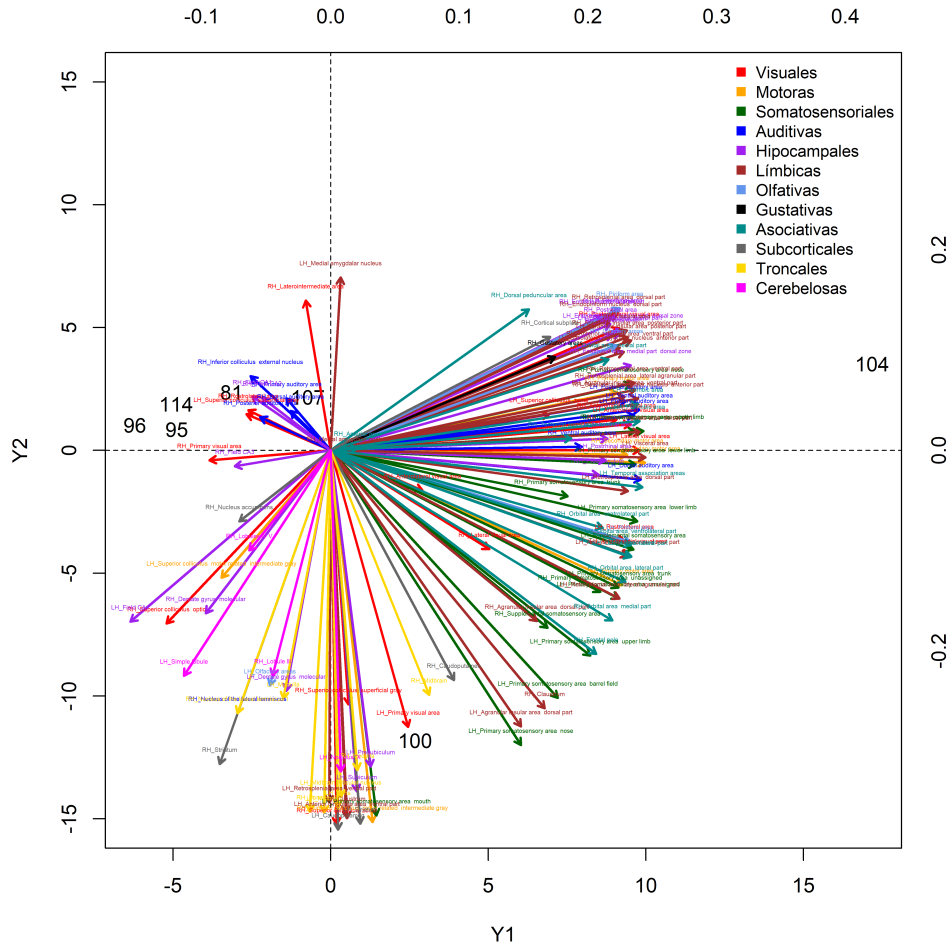


Figura 3.6. Gráfico biplot para las dos primeras componentes principales, mostrando a los individuos a través de sus respectivas etiquetas numéricas y a las variables (regiones cerebrales) mediante vectores. A su vez, los colores permiten distinguir el área funcional al que pertenece cada variable, según la leyenda.

En primer lugar, el ángulo formado entre dos vectores refleja la correlación lineal entre las variables que representan. Un ángulo cercano a 0° indica una fuerte correlación positiva entre ellas, uno próximo a 90° refleja ausencia de correlación y un ángulo cercano a 180° evidencia una fuerte correlación negativa. Así, la presencia de ángulos pequeños entre gran parte de los vectores indica una fuerte correlación positiva entre la mayoría de las regiones cerebrales, tal y como vimos en la matriz de correlaciones R (véase Figura 3.4). De hecho, cabe destacar que, en general, las estructuras pertenecientes al mismo área funcional están fuertemente correlacionadas.

Asimismo, se puede observar que el grupo de regiones situadas en el semieje negativo presenta un ángulo cercano a 90° , con gran parte de las regiones concentradas en el semiplano derecho $\{(Y_1, Y_2) \in \mathbb{R}^2 \mid Y_1 > 0\}$, lo que refleja una baja correlación de este grupo con las demás estructuras. En particular, dentro de este último grupo cabe destacar la presencia de varias estructuras del tronco encefálico.

Con respecto a las observaciones, vemos en primer lugar que el individuo 104 está muy influenciado por la primera componente principal, al igual que ocurre, en menor medida, con los individuos 95, 96 y 114. Por el contrario, el individuo 100 está fuertemente vin-

culado a la segunda componente principal. Asimismo, cabe destacar que los sujetos 95, 96 y 114 se sitúan próximos entre sí, lo que indica que comparten características similares.

Por otro lado, este gráfico también permite visualizar la contribución de las estructuras cerebrales en las dos primeras componentes, cuyos valores (*loadings*), correspondientes a los coeficientes de las combinaciones lineales que definen cada componente (véase Definición 2.15), se recogen en la Tabla B.2 del Apéndice B, permitiendo así analizar las regiones con mayor influencia en dichas componentes.

En primer lugar, la primera componente principal captura la mayor parte de la varianza conjunta de regiones correlacionadas. En particular, está dominada principalmente por gran parte de las regiones somatosensoriales, motoras y asociativas. Asimismo, se observan contribuciones positivas destacables en algunas regiones límbicas como RH_Anterior cingulate area, dorsal part y LH_Anterior cingulate area, dorsal part, así como en las áreas hipocampales LH_Perirhinal area, RH_Perirhinal area, RH_Entorhinal area, medial part, dorsal zone y LH_Entorhinal area, medial part, dorsal zone y en las regiones auditivas LH_Primary auditory area, LH_Dorsal auditory area y LH_Ventral auditory area.

Con respecto a la segunda componente principal, las mayores contribuciones en valor absoluto corresponden principalmente a las estructuras troncales RH_Medulla, RH_Pons, LH_Pons, LH_Midbrain y LH_Midbrain reticular nucleus, junto con las regiones límbicas LH_Anterior cingulate area, ventral part, LH_Clastrum, LH_Lateral septal nucleus, rostral part, LH_Retrosplenial area, ventral part y LH-Taenia tecta, dorsal part, las regiones subcorticales LH_Caudoputamen, LH_Striatum y RH_Striatum, la región cerebelosa LH_Nodulus(X), la región somatosensorial LH_Primary somatosensory area, mouth, las regiones hipocampales LH_Presubiculum y LH_Subiculum, la región visual RH_Superior colliculus, motor related, intermediate gray y finalmente la región motora RH_Superior colliculus, zonal.

En conclusión, el PCA permitió condensar la información del conjunto de variables original en las tres primeras componentes principales, capaces de explicar más del 83 % de la variabilidad total de los datos, concentrada especialmente en la primera componente, con más del 50 %, y la segunda, con alrededor del 22 %. La primera componente principal está fuertemente correlacionada con las regiones de las áreas motora, asociativa y somatosensorial, así como con algunas estructuras límbicas, auditivas e hipocampales. Por otro lado, la segunda componente principal está fuertemente correlacionada con varias regiones troncales, límbicas, hipocampales y subcorticales, observándose además, en menor medida, la contribución aislada de una región visual, una somatosensorial y una motora. Por tanto, se observa una clara diferenciación entre ambas componentes, donde la primera está dominada principalmente por la mayoría de regiones corticales, y la segunda está asociada fundamentalmente a estructuras localizadas debajo de la corteza cerebral. Sin embargo, estos resultados deben interpretarse con cautela, ya que la alta dimensionalidad frente a un tamaño muestral pequeño compromete la estabilidad del PCA y lo hace sensible a pequeñas variaciones en los datos, por lo que las conclusiones deben considerarse de carácter exploratorio.

3.4.2. Aplicación del Test T^2 de Hotelling

Una vez obtenidas las tres primeras componentes principales, se considera el vector aleatorio tridimensional $\mathbf{Y} = (Y_1, Y_2, Y_3)^\top$, para el que se dispone de dos muestras aleatorias e independientes correspondientes a los grupos control y privado, de tamaños 4 y 3, respectivamente.

El objetivo de este apartado será evaluar si existen diferencias significativas entre los vectores de medias de ambos grupos, para lo que se considera el siguiente contraste:

$$\begin{cases} H_0 : \boldsymbol{\mu}^{(C)} = \boldsymbol{\mu}^{(D)} \\ H_a : \boldsymbol{\mu}^{(C)} \neq \boldsymbol{\mu}^{(D)}, \end{cases}$$

siendo $\boldsymbol{\mu}^{(C)} = (\mu_1^{(C)}, \mu_2^{(C)}, \mu_3^{(C)})^\top$ el vector de medias del grupo control y siendo $\boldsymbol{\mu}^{(D)} = (\mu_1^{(D)}, \mu_2^{(D)}, \mu_3^{(D)})^\top$ el vector de medias del grupo privado.

Para ello, se emplea el test T^2 de Hotelling, el cual requiere el supuesto de normalidad en ambos grupos para garantizar la validez del contraste. Sin embargo, al disponer de dos muestras tan pequeñas, cualquier aproximación normal resulta poco fiable, independientemente de lo que indique el test de normalidad de Shapiro-Wilk. Como solución se emplea el test de permutaciones basado en el estadístico T^2 de Hotelling (véase Teorema 2.18), considerando un rango amplio de permutaciones, desde $B = 100$ hasta $B = 7!$. Los resultados obtenidos se recogen en la Tabla 3.19.

B	Estadístico	p-valor
100	0.8541	0.5501
500	0.8541	0.5620
1000	0.8541	0.5940
5040	0.8541	0.5629

Tabla 3.19. Resultados del test de permutaciones basado en el estadístico T^2 de Hotelling, con $B = 100$, $B = 500$, $B = 1000$ y $B = 7!$.

De este modo, se obtiene el estadístico de contraste $\tilde{T}_{4,3}^2 = 0.8541$, junto con los p -valores del test para los diferentes casos considerados, los cuales resultan muy similares entre sí y se sitúan por encima de los niveles de significación habituales (1% y 5%). En consecuencia, no se tienen evidencias significativas en contra de la hipótesis nula, lo que sugiere que los vectores de medias de ambos grupos son iguales. Por tanto, la proporción de células relativo a la zona de inyección en las regiones de la corteza motora, asociativa y somatosensorial (principales contribuyentes de la primera componente principal), así como gran parte de las regiones troncales, límbicas, hipocampales y subcorticales (estructuras adicionales con mayor peso en la segunda componente principal) resulta comparable entre los grupos control y privado. No obstante, el reducido tamaño muestral puede afectar a la estabilidad del test y aumentar su sensibilidad a variaciones en los datos.

Capítulo 4

Conclusiones y líneas futuras

En este último capítulo se recogen los resultados y conclusiones obtenidas a lo largo del proyecto. Además, se proponen futuras líneas de investigación y mejoras metodológicas que podrían dar continuidad a este trabajo.

4.1. Conclusiones

El objetivo de este Trabajo Fin de Máster fue aplicar diferentes herramientas estadísticas que permitan al laboratorio *Brain Plasticity* del CiMUS profundizar en el estudio de la plasticidad cerebral en la edad adulta, empleando para ello un modelo experimental de privación sensorial en ratones. En concreto, se realizó una inyección viral en el *barrel field* derecho de nueve ratones, cuatro de ellos privados, con el objetivo de analizar si dicha área fue capaz de procesar información proveniente de otras áreas sensoriales. Para ello, el experimento se llevó a cabo en dos etapas.

En la primera fase del estudio se evaluó la comparabilidad de las inyecciones realizadas en la región *barrel field* de los distintos animales mediante dos enfoques complementarios. Desde una perspectiva cuantitativa, los modelos lineales generalizados mixtos permitieron analizar la distribución del número de neuronas entre las distintas capas corticales del *barrel field*, teniendo en cuenta la estructura de medidas repetidas presente en los datos. Por otro lado, desde el segundo enfoque, los procesos puntuales permitieron estudiar la distribución espacial de las neuronas dentro de la región de inyección en los distintos animales. Los resultados de ambos análisis respaldaron la homogeneidad de las inyecciones, a excepción de los individuos 82 y 105, que fueron excluidos en la siguiente etapa del experimento.

Una vez confirmada la homogeneidad de las inyecciones entre los siete ratones restantes, se procedió al estudio de la conectividad a escala de todo el cerebro a partir de las proporciones de células relativas a la zona de inyección. El análisis de componentes principales permitió reducir la dimensionalidad del conjunto de datos y sintetizar gran mayor parte de la información en las tres primeras componentes principales, estando la primera dominada por regiones corticales motoras, asociativas y somatosensoriales, y la segunda por estructuras más profundas como las regiones troncales, límbicas e hipocámpales. Seguidamente, se llevó a cabo la comparación multivariante de las medias entre los grupos control y privado mediante el test de permutaciones basado en el estadístico T^2 de Hotelling, el cual no proporcionó evidencias suficientes para rechazar la hipótesis de igualdad de medias entre ambos grupos.

Por tanto, se observaron proporciones de células similares entre los grupos control y privado en las regiones corticales motoras, asociativas y somatosensoriales, así como en diversas estructuras troncales, límbicas, hipocampales y subcorticales, lo que sugiere que el *barrel field* derecho de los individuos privados fue capaz de procesar información proveniente de estas áreas.

Finalmente, señalar que el tamaño muestral reducido constituye una limitación general del estudio, ya que puede afectar a la potencia estadística de los análisis realizados, incrementar la variabilidad interindividual y puede condicionar la robustez de las conclusiones, por lo que estas deben considerarse de carácter exploratorio.

4.2. Líneas futuras

A raíz del desarrollo de este Trabajo de Fin de Máster, se han detectado algunas líneas de investigación que podrían abordarse en el futuro.

En primer lugar, sería conveniente ampliar el número de ratones analizados, permitiendo obtener estimaciones más robustas y mejorar la capacidad para detectar posibles cambios asociados a la privación sensorial, si bien esto supondría un aumento considerable tanto de los costes como de la complejidad experimental. Del mismo modo, resultaría de interés realizar más mediciones a lo largo del tiempo, con el fin de analizar mejor cómo evoluciona la reorganización cerebral y determinar cuándo aparecen los cambios en la conectividad.

Asimismo, desde el punto de vista metodológico, sería interesante extender el análisis mediante técnicas basadas en la teoría de grafos, representando las neuronas como nodos y las conexiones entre ellas como aristas. Este enfoque permitiría estudiar la organización global de la conectividad cerebral mediante medidas como la centralidad, la modularidad o la eficiencia de la red, aportando una perspectiva complementaria al análisis cuantitativo realizado.

Apéndice A

Bases de datos del experimento

Este apéndice ofrece una descripción detallada de las mediciones celulares recogidas en los dos tipos de bases de datos empleadas en este trabajo: la base de recuentos celulares y las bases de coordenadas celulares.

En primer lugar, se describe el conjunto de atributos que conforman la base de datos de los recuentos celulares.

Base de datos de los recuentos celulares

- **id**: identificador numérico único asignado a cada estructura cerebral.
- **name**: nombre anatómico de la región cerebral, tracto o fibra nerviosa. En el caso de regiones corticales se incluye la capa (*layer*) si está especificada.
- **volume**: número de vóxeles¹ que componen a cada estructura cerebral, con valores comprendidos entre 4 218 750 y 13 010 187 500 vóxeles.
- **hemisphere**: hemisferio cerebral al que pertenece la región, codificado como 0 para el hemisferio derecho y 255 para el izquierdo.
- **Controls_counts_i**, con $i \in \{81, 82, 95, 96, 100\}$: número de células observadas por región en los individuos control 81, 82, 95, 96 ó 100.
- **mean_Controls**: número medio de células por región cerebral (fila) calculada para los ratones del grupo control (columnas del 5 al 9).
- **sd_Controls**: desviación típica del número de células por región cerebral en el grupo control.
- **Deprived_counts_i**, con $i \in \{104, 105, 107, 114\}$: número de células observadas por región en los individuos deprivados 104, 105, 107 ó 114.
- **mean_Deprived**: número medio de células por región cerebral (fila) calculada para los ratones del grupo deprivado (columnas del 12 al 15).
- **sd_Deprived**: desviación típica del número de células por región cerebral en el grupo deprivado.

¹El vóxel es la unidad mínima de volumen que compone un objeto tridimensional. En nuestro caso se emplean vóxeles de $1,62 \mu\text{m} \times 1,62 \mu\text{m} \times 3 \mu\text{m}$.

- **p_value**: p-valor obtenido al contrastar la hipótesis nula de igualdad en el número de células entre los grupos control y privado, realizado de manera independiente para cada región cerebral (fila) mediante la prueba t de Student.

En particular, para facilitar la comprensión de esta base de datos, se muestran en la Tabla A.1 las cinco primeras filas.

id	name	volume	hemisphere	Controls_counts_81	Controls_counts_82
0	universe	348622000000	0	105	1224
2	Primary somatosensory area, mouth, layer 6b	57343750	0	1	1
6	internal capsule	1049500000	0	0	50
7	Principal sensory nucleus of the trigeminal	551281250	0	0	0
9	Primary somatosensory area, trunk, layer 6a	116234375	0	33	704

id	Controls_counts_95	Controls_counts_96	Controls_counts_100	mean_Controls	sd_Controls	Deprived_counts_104
0	310	398	841	575.6	451.2231	747
2	9	2	2	3.0	3.3912	51
6	77	108	58	58.6	39.6459	80
7	0	1	1	0.4	0.5477	1
9	140	66	147	218.0	275.9755	350

id	Deprived_counts_105	Deprived_counts_107	Deprived_counts_114	mean_Deprived	sd_Deprived	p_value
0	746	167	140	450.00	342.5463	0.6496
2	3	3	1	14.50	24.3926	0.4166
6	131	9	22	60.50	56.2287	0.9565
7	0	2	1	1.00	0.8165	0.2626
9	42	55	227	168.50	147.4777	0.7418

Tabla A.1. Extracto de las cinco primeras filas de la base de datos de los recuentos celulares.

Teniendo en cuenta el preprocesado de datos descrito en la Sección 3.1, eliminaremos las filas 1 (**name** = *universe*) y 3 (**name** = *internal capsule*) de la Tabla A.1, por tratarse esta última de un tracto de fibras. Asimismo, en la cuarta fila (región *Principal sensory nucleus of the trigeminal*) no se verifica que todos los recuentos celulares sean estrictamente mayor que cero, por lo que también será eliminada. De este modo, este extracto queda reducido únicamente a dos filas, la segunda y la quinta.

Por otro lado, en la Tabla A.2 se muestra un extracto de la base de datos modificada (es decir, tras completar las 3 etapas del preprocesado de datos), correspondiente a las tres primeras regiones, junto con el grupo experimental al que pertenece cada individuo.

id	RH.Agranular insular area, dorsal part	LH.Agranular insular area, dorsal part	RH.Agranular insular area, posterior part	...	Grupo
81	62	7	22	...	Control
82	88	12	331	...	Control
95	87	9	47	...	Control
96	166	17	35	...	Control
100	91	18	16	...	Control
104	322	44	487	...	Deprived
105	347	22	112	...	Deprived
107	71	8	37	...	Deprived
114	65	4	16	...	Deprived

Tabla A.2. Extracto de las 3 primeras variables (columnas) de la base de datos modificada.

Seguidamente, se ofrece una descripción detallada del conjunto de atributos que conforman las bases de datos de las coordenadas celulares correspondientes a los 9 individuos.

Base de datos de las coordenadas celulares

- **x**: coordenada que representa la posición horizontal de la célula en la región cerebral.
- **y**: coordenada que representa la posición vertical de la célula en la región cerebral.
- **z**: coordenada que representa la profundidad de la célula en la región cerebral.
- **size**: tamaño estimado de la célula, calculado a partir del número de vóxeles que la componen.
- **x_t**: coordenada x transformada o registrada en el atlas de referencia.
- **y_t**: coordenada y transformada o registrada en el atlas de referencia.
- **z_t**: coordenada z transformada o registrada en el atlas de referencia.
- **order**: identificador numérico único asignado a cada célula.
- **name**: nombre anatómico de la región cerebral a la que pertenece la célula, incluyendo la capa si está especificada.

En particular, para facilitar la comprensión de este conjunto de datos, se muestran a continuación las seis primeras filas para el individuo 81.

x	y	z	size	ssource	xt	yt	zt	order	name
3215	3190	63	152	9840723	253.95	314.54	10.74	332	Retrosplenial area, dorsal part, layer 1
2758	3015	67	56	15002249	218.10	301.21	16.35	332	Retrosplenial area, dorsal part, layer 1
2058	2815	70	118	18837674	162.84	287.00	24.15	385	No label
2499	3146	71	85	20225798	197.91	312.41	17.19	333	Retrosplenial area, dorsal part, layer 2/3
2404	2583	78	187	6006774	190.63	268.06	25.23	306	Retrosplenial area, lateral agranular part, layer 2/3
2740	2472	78	112	6061342	216.62	258.76	24.25	332	Retrosplenial area, dorsal part, layer 1

Tabla A.3. Extracto de las seis primeras filas de la base de datos de las coordenadas celulares para el individuo 81.

Apéndice B

Tablas relativas al PCA

En primer lugar, en la Tabla B.1 se muestra la agrupación de las regiones cerebrales según su área funcional.

Área	Regiones cerebrales	Área	Regiones cerebrales
Visual	RH.Anterolateral visual area LH.Anterolateral visual area RH.Anteromedial visual area LH.Anteromedial visual area RH.Lateral visual area LH.Lateral visual area RH.Posterolateral visual area RH.Primary visual area LH.Primary visual area RH.Superior colliculus, optic RH.Superior colliculus, superficial gray LH.Superior colliculus, superficial gray RH.Superior colliculus, zonal LH.Superior colliculus, zonal RH.Laterointermediate area RH.Rostrolateral area LH.Rostrolateral area	Motora	RH.Primary motor area LH.Primary motor area RH.Secondary motor area LH.Secondary motor area RH.Superior colliculus, motor related, intermediate gray LH.Superior colliculus, motor related, intermediate gray
Auditiva	RH.Primary auditory area LH.Primary auditory area RH.Dorsal auditory area LH.Dorsal auditory area RH.Posterior auditory area LH.Posterior auditory area RH.Ventral auditory area LH.Ventral auditory area RH.Inferior colliculus, external nucleus RH.Nucleus of the lateral lemniscus	Asociativa	RH.Prelimbic area LH.Prelimbic area RH.Orbital area, lateral part LH.Orbital area, lateral part RH.Orbital area, medial part LH.Orbital area, medial part RH.Orbital area, ventrolateral part LH.Orbital area, ventrolateral part RH.Frontal pole RH.Temporal association areas LH.Temporal association areas RH.Anterior area LH.Anterior area RH.Dorsal peduncular area
Olfativa	RH.Olfactory areas LH.Olfactory areas RH.Piriform area LH.Piriform area	Gustativa	RH.Gustatory areas

Área	Regiones cerebrales	Área	Regiones cerebrales
Límbica	RH.Agranular insular area, dorsal part LH.Agranular insular area, dorsal part RH.Agranular insular area, posterior part LH.Agranular insular area, posterior part RH.Agranular insular area, ventral part RH.Anterior cingulate area, dorsal part LH.Anterior cingulate area, dorsal part RH.Anterior cingulate area, ventral part LH.Anterior cingulate area, ventral part RH.Retrosplenial area, dorsal part LH.Retrosplenial area, dorsal part RH.Retrosplenial area, lateral agranular part LH.Retrosplenial area, lateral agranular part RH.Retrosplenial area, ventral part LH.Retrosplenial area, ventral part RH.Visceral area LH.Visceral area RH.Medial amygdalar nucleus LH.Medial amygdalar nucleus RH.Basolateral amygdalar nucleus, anterior part LH.Basolateral amygdalar nucleus, anterior part RH.Endopiriform nucleus, dorsal part RH.Clastrum LH.Clastrum	Subcortical	RH.Caudoputamen LH.Caudoputamen RH.Striatum LH.Striatum RH.Nucleus accumbens RH.Cortical subplate
Somatosensorial	RH.Supplemental somatosensory area LH.Supplemental somatosensory area RH.Primary somatosensory area, upper limb LH.Primary somatosensory area, upper limb RH.Primary somatosensory area, lower limb LH.Primary somatosensory area, lower limb RH.Primary somatosensory area, trunk LH.Primary somatosensory area, trunk RH.Primary somatosensory area, mouth LH.Primary somatosensory area, mouth RH.Primary somatosensory area, nose LH.Primary somatosensory area, nose RH.Primary somatosensory area, unassigned LH.Primary somatosensory area, unassigned RH.Primary somatosensory area, barrel field LH.Primary somatosensory area, barrel field	Hipocampal	RH.Field CA1 LH.Field CA1 RH.Field CA2 RH.Field CA3 RH.Dentate gyrus, molecular LH.Dentate gyrus, molecular RH.Presubiculum LH.Presubiculum RH.Parasubiculum RH.Entorhinal area, lateral part LH.Entorhinal area, lateral part RH.Entorhinal area, medial part, dorsal zone LH.Entorhinal area, medial part, dorsal zone RH.Perirhinal area LH.Perirhinal area RH.Postrhinal area LH.Postrhinal area RH.Ectorhinal area/ LH.Ectorhinal area/ LH.Subiculum

Área	Regiones cerebrales	Área	Regiones cerebrales
Troncal	RH.Midbrain LH.Midbrain LH.Midbrain reticular nucleus RH.Medulla LH.Medulla RH.Pons LH.Pons RH.Nucleus of the lateral lemniscus	Cerebelo	RH.Lobule III RH.Lobules IV-V LH.Simple lobule LH.Nodulus (X)

Tabla B.1. Agrupación de las regiones cerebrales según su área funcional.

Por otro lado, en la Tabla B.2 se muestran las contribuciones de las regiones cerebrales (variables) sobre las dos primeras componentes principales.

Región cerebral	Y ₁	Y ₂
RH.Agranular insular area, dorsal part	0.0801	-0.0829
LH.Agranular insular area, dorsal part	0.0744	-0.1344
RH.Agranular insular area, posterior part	0.1146	0.0572
LH.Agranular insular area, posterior part	0.1161	0.0551
RH.Agranular insular area, ventral part	0.1135	0.0291
RH.Anterior area	0.0103	0.0040
LH.Anterior area	0.1167	-0.0495
RH.Anterior cingulate area, dorsal part	0.1199	0.0098
LH.Anterior cingulate area, dorsal part	0.1154	-0.0504
RH.Anterior cingulate area, ventral part	0.1118	0.0516
LH.Anterior cingulate area, ventral part	0.0071	-0.1778
RH.Anterior cingulate area, ventral part, 6a	0.0316	-0.0369
RH.Anterolateral visual area	0.0354	-0.0166
LH.Anterolateral visual area	0.1158	0.0158
RH.Anteromedial visual area	0.1111	0.0148
LH.Anteromedial visual area	0.1156	-0.0490
RH.Basolateral amygdalar nucleus, anterior part	0.1188	0.0277
LH.Basolateral amygdalar nucleus, anterior part	0.1131	0.0506
RH.Caudoputamen	0.0482	-0.1098
LH.Caudoputamen	0.0039	-0.1846
RH.Clastrum	0.0839	-0.1263
LH.Clastrum	0.0072	-0.1778
RH.Cortical subplate	0.0853	0.0525

Región cerebral	Y ₁	Y ₂
RH.Dentate gyrus, molecular	-0.0489	-0.0750
LH.Dentate gyrus, molecular	-0.0172	-0.1149
RH.Dorsal auditory area	-0.0159	0.0163
LH.Dorsal auditory area	0.1200	-0.0116
RH.Dorsal peduncular area	0.0772	0.0656
RH.Ectorhinal area/	0.1090	0.0556
LH.Ectorhinal area/	0.1077	0.0048
RH.Endopiriform nucleus, dorsal part	0.1082	0.0663
RH.Entorhinal area, lateral part	0.1051	0.0650
LH.Entorhinal area, lateral part	0.1118	0.0590
RH.Entorhinal area, medial part, dorsal zone	0.1160	0.0428
LH.Entorhinal area, medial part, dorsal zone	0.1128	0.0611
RH.Field CA1	-0.0314	0.0267
LH.Field CA1	-0.0777	-0.0830
RH.Field CA2	-0.0274	0.0253
RH.Field CA3	-0.0378	-0.0081
RH.Frontal pole	0.1033	-0.0976
RH.Gustatory areas	0.0863	0.0477
RH.Inferior colliculus, external nucleus	-0.0322	0.0374
LH.Lateral septal nucleus, rostral part	-0.0014	-0.1661
RH.Lateral visual area	0.0624	-0.0529
LH.Lateral visual area	0.1196	0.0024
RH.Laterointermediate area	-0.0097	0.0707
RH.Lobule III	-0.0223	-0.1113
RH.Lobules IV-V	-0.0307	-0.0552
RH.Medial amygdalar nucleus	0.0058	0.0035
LH.Medial amygdalar nucleus	0.0041	0.0797
RH.Medulla	0.0118	-0.1573
LH.Medulla	-0.0168	-0.1261
RH.Midbrain	0.0387	-0.1197
LH.Midbrain	-0.0013	-0.1778
LH.Midbrain reticular nucleus	0.0053	-0.1711

Región cerebral	Y ₁	Y ₂
LH.Nodulus (X)	0.0052	-0.1584
RH.Nucleus accumbens	-0.0348	-0.0373
RH.Nucleus of the lateral lemniscus	-0.0355	-0.1253
RH.Olfactory areas	0.1102	0.0542
LH.Olfactory areas	-0.0228	-0.1157
RH.Orbital area, lateral part	0.1145	-0.0621
LH.Orbital area, lateral part	0.1164	-0.0500
RH.Orbital area, medial part	0.1096	-0.0811
LH.Orbital area, medial part	0.1074	0.0453
RH.Orbital area, ventrolateral part	0.1057	-0.0359
LH.Orbital area, ventrolateral part	0.1164	-0.0444
RH.Parasubiculum	0.1113	0.0594
RH.Perirhinal area	0.1076	0.0650
LH.Perirhinal area	0.1105	0.0675
RH.Piriform area	0.1117	0.0702
LH.Piriform area	0.1162	-0.0437
RH.Pons	-0.0069	-0.1759
LH.Pons	0.0041	-0.1760
RH.Posterior auditory area	-0.0275	0.0172
LH.Posterior auditory area	0.1139	0.0264
RH.Posterolateral visual area	0.1112	0.0619
RH.Postrhinal area	0.1080	0.0625
LH.Postrhinal area	0.1071	-0.0060
RH.Prelimbic area	0.1194	0.0167
LH.Prelimbic area	0.1188	0.0249
RH.Presubiculum	0.1039	-0.0127
LH.Presubiculum	0.0166	-0.1562
RH.Primary auditory area	-0.0170	0.0196
LH.Primary auditory area	0.1171	0.0197
RH.Primary motor area	0.1111	0.0271
LH.Primary motor area	0.1131	-0.0635
LH.Primary somatosensory area, barrel field	0.0887	-0.1184

Región cerebral	Y ₁	Y ₂
RH.Primary somatosensory area, lower limb	0.1179	-0.0038
LH.Primary somatosensory area, lower limb	0.1189	-0.0325
RH.Primary somatosensory area, mouth	0.1177	0.0116
LH.Primary somatosensory area, mouth	0.0186	-0.1762
RH.Primary somatosensory area, nose	0.1161	0.0346
LH.Primary somatosensory area, nose	0.0744	-0.1411
RH.Primary somatosensory area, trunk	0.0918	-0.0181
LH.Primary somatosensory area, trunk	0.1118	-0.0645
RH.Primary somatosensory area, unassigned	0.1050	-0.0668
LH.Primary somatosensory area, unassigned	0.1119	-0.0700
RH.Primary somatosensory area, upper limb	0.1210	0.0116
LH.Primary somatosensory area, upper limb	0.1011	-0.0979
RH.Primary visual area	-0.0477	-0.0040
LH.Primary visual area	0.0303	-0.1320
RH.Retrosplenial area, dorsal part	0.1113	0.0684
LH.Retrosplenial area, dorsal part	0.1150	-0.0173
RH.Retrosplenial area, lateral agranular part	0.1177	0.0321
LH.Retrosplenial area, lateral agranular part	0.1122	-0.0702
RH.Retrosplenial area, ventral part	0.1175	0.0341
LH.Retrosplenial area, ventral part	0.0000	-0.1696
RH.Rostrolateral area	-0.0330	0.0207
LH.Rostrolateral area	0.1138	-0.0408
RH.Secondary motor area	0.1153	-0.0032
LH.Secondary motor area	0.1218	-0.0042
LH.Simple lobule	-0.0568	-0.1083
RH.Striatum	-0.0424	-0.1541
LH.Striatum	0.0127	-0.1824
LH.Subiculum	0.0107	-0.1644
RH.Superior colliculus, motor related, intermediate gray	0.0171	-0.1806
LH.Superior colliculus, motor related, intermediate gray	-0.0427	-0.0620
RH.Superior colliculus, optic	-0.0635	-0.0861
RH.Superior colliculus, superficial gray	0.0081	-0.1253

Región cerebral	Y_1	Y_2
LH.Superior colliculus, superficial gray	-0.0335	0.0179
RH.Superior colliculus, zonal	0.0033	-0.1826
LH.Superior colliculus, zonal	0.0836	0.0210
RH.Supplemental somatosensory area	0.0849	-0.0850
LH.Supplemental somatosensory area	0.1175	-0.0460
RH.Taenia tecta, dorsal part	0.0751	0.0039
LH.Taenia tecta, dorsal part	0.0134	-0.1827
RH.Temporal association areas	0.0939	0.0028
LH.Temporal association areas	0.1209	-0.0164
RH.Ventral auditory area	0.0977	0.0012
LH.Ventral auditory area	0.1193	0.0221
RH.Visceral area	0.1145	0.0595
LH.Visceral area	0.1219	-0.0020

Tabla B.2. Contribuciones de las variables sobre las componentes principales Y_1 e Y_2 .

Bibliografía

- [1] Agresti, A. (1996). *An introduction to categorical data analysis*. Wiley.
- [2] Anderson, T. W. (1994). *An Introduction to Multivariate Statistical Analysis*. Wiley.
- [3] Armstrong, R. A. (2014). *When to use the Bonferroni correction*. *Ophthalmic Physiol Optics*, 34(5): 502-508. <https://doi.org/10.1111/opo.12131>.
- [4] Bach, L., Al-muhtasib, N., King, L. y Thometz, N. (2025). *Human Physiology*. An Open Educational Resource OER for Social Justice, Chapter 5.
- [5] Baddeley, A. y Turner, R. (2005). Spatstat: an R package for analyzing spatial point patterns. *Journal of Statistical Software*, 12(6): 1–42. <https://dx.doi.org/10.18637/jss.v012.i06>.
- [6] Bazira, P. (2021). *An overview of the nervous system*. *Surgery (Oxford)*, 39(8): 451-462. <https://doi.org/10.1016/j.mpsur.2021.06.012>.
- [7] Bekolay, T. (2011). *Learning in large-scale spiking neural networks*. <https://api.semanticscholar.org/CorpusID:16037628>
- [8] Chacón, J. E. y Duong, T. (2010). *Multivariate plug-in bandwidth selection with unconstrained pilot bandwidth matrices*. *TEST*, 19(2): 375–398. <https://dx.doi.org/10.1007/s11749-009-0168-4>.
- [9] Chung, E. y Romano, J. P. (2016). *Multivariate and multiple permutation tests*. *Journal of Econometrics*, 193(1): 76-91. <https://doi.org/10.1016/j.jeconom.2016.01.003>.
- [10] Cucala, L. (2006). *Espacements bidimensionnels et données entachées deerreurs dans lanalyse des processus ponctuels spatiaux*. Tesis Doctoral, Université des Sciences Sociales, Toulouse I.
- [11] Crujeiras Casais, R. y Conde Amboague, M. (2024). Regresión generalizada y modelos mixtos. Material docente del Máster en Técnicas Estadísticas, Universidade de Santiago de Compostela.
- [12] Denk, W. y Horstmann, H. (2004). *Serial block-face scanning electron microscopy to reconstruct three-dimensional tissue nanostructure*. *PLoS Biology*, 2(11). <https://doi.org/10.1371/journal.pbio.0020329>.
- [13] Diggle, P. (2013). *Statistical Analysis of Spatial and Spatio-Temporal Point Patterns*. Chapman and Hall/CRC.
- [14] Dunteman, G. H. (1989). *Principal components analysis*. Newbury Park.
- [15] Fuchsl, E. y Flügge, G. (2014). *Adult neuroplasticity: more than 40 years of research*. *Neural Plasticity*.

- [16] Fuentes-Santos, I., González-Manteiga, W. y Mateu, J. (2016). *Consistent smooth bootstrap Kernel intensity estimation for inhomogeneous spatial Poisson point processes*. Scandinavian Journal of Statistics, 43(2): 416–435. <https://dx.doi.org/10.1111/sjos.12183>.
- [17] Fuentes-Santos, I., González-Manteiga, W. y Mateu, J. (2017). *A nonparametric test for the comparison of first-order structures of spatial point processes*. Spatial Statistics, 22(2): 240-260. <https://doi.org/10.1016/j.spasta.2017.02.007>.
- [18] Fuentes-Santos, I., González-Manteiga, W. y Mateu, J. (2023). *Testing similarity between first-order intensities of spatial point processes. A comparative study*. Communications in Statistics - Simulation and Computation, 52(5): 2210-2230. <https://doi.org/10.1080/03610918.2021.1901118>.
- [19] Gazerani, P. (2025). *The neuroplastic brain: current breakthroughs and emerging frontiers*. Brain Research, 1858. <https://doi.org/10.1016/j.brainres.2025.149643>.
- [20] Guo, C., Peng, J., Zhang, Y., Li, A., Li, Y., Yuan, J., Feng, X., Ren, M., Gong, H. y Chen, S. (2017). *Single-axon level morphological analysis of corticofugal projection neurons in mouse barrel field*. Scientific Reports, 7(2846). <https://doi.org/10.1038/s41598-017-03000-8>.
- [21] Haines, D., Ard, M.D. y Schenck, M.P. (2006). *Principios de neurociencia*. Elsevier.
- [22] Kéry, M. (2010). *Poisson Mixed-Effects Model (Poisson GLMM)*. Introduction to WinBUGS for Ecologists, Academic Press, 203-209. <https://doi.org/10.1016/B978-0-12-378605-0.00016-8>.
- [23] Kirst, C., Vieites-Prado, A., Topilko, T., Bertin, P., Gerschenfeld, G., Verny, F., Topilko, P., Michalski, N., Tessier-Lavigne, M. y Renier, N. (2020). *Mapping the Fine-Scale Organization and Plasticity of the Brain Vasculature*. Cell, 180(4): 780-795. <https://doi.org/10.1016/j.cell.2020.01.028>.
- [24] Krzanowski, W. J. (1988). *Principles of Multivariate Analysis. A User's Perspective*. Oxford University Press.
- [25] Larkum, M., Petro, L., Sachdev, R. y Muckli, L. (2018). *A Perspective on Cortical Layering and Layer-Spanning Neuronal Elements*. Front Neuroanat. <https://doi.org/10.3389/fnana.2018.00056>.
- [26] Lv, J., Guo, L., Hu, X., Zhang, T., Li, K., Zhang, D., Yang, J. y Liu, T. (2010). *Fiber-centered analysis of brain connectivities using DTI and resting state FMRI data*. Medical Image Computing and Computer-Assisted Intervention, 13(Pt 2): 143-150.
- [27] Lix, L. M. y Keselman, H. J. (2004). *Multivariate tests of means in independent groups designs. Effects of covariance heterogeneity and nonnormality*. Evaluation & the Health Professions, 27(1): 45-69.
- [28] Lydersen, S. (2024). *Adjustment of p values for multiple hypotheses: why, when and how*. Annals of the Rheumatic Diseases, 83(10): 1254-1255. <https://doi.org/10.1136/ard-2024-225537>.
- [29] Ma, Q., Dunton, G. y Hedeker, D. (2025). *Negative binomial mixed effects location-scale models for intensive longitudinal count-type physical activity data provided by wearable devices*. Biometrics, 81(3). <https://doi.org/10.1093/biomtc/ujaf099>.

- [30] Maćkiewicz, A. y Ratajczak, W. (1993). *Principal components analysis (PCA)*. Computers & Geosciences, 19(3): 303-342. [https://doi.org/10.1016/0098-3004\(93\)90090-R](https://doi.org/10.1016/0098-3004(93)90090-R).
- [31] Mateu, J. y Jalilian, A. (2022). *Spatial point processes and neural networks: A convenient couple*. Spatial Statistics, 50. <https://doi.org/10.1016/j.spasta.2022.100644>.
- [32] Mason, R.L. y Young, J.C. (2011). *Hotelling's T^2 Statistic*. International Encyclopedia of Statistical Science, 638-640. Springer.
- [33] Mowery, T. M., y Garraghty, P. E. (2023). *Adult neuroplasticity employs developmental mechanisms*. Frontiers in Systems Neuroscience.
- [34] McCullough, P. y Nelder, J.A. (1989). *Generalized Linear Models*, 2nd ed. Chapman and Hall/CRC.
- [35] Nelder, J. A. y Wedderburn, R. W. M. (1972). *Generalized Linear Models*. Journal of the Royal Statistical Society. Series A (General), 135(3): 370-384. <https://doi.org/10.2307/2344614>.
- [36] Perry, G., Miller, B. y Enright, N. (2006). *A comparison of methods for the statistical analysis of spatial point patterns in plant ecology*. Plant Ecol, 187: 59-82. <https://doi.org/10.1007/s11258-006-9133-4>.
- [37] Renier, N., Wu, Z., Simon, D., Yang, J., Ariel, P. y Tessier-Lavigne, M. (2014). *iDISCO: a simple, rapid method to immunolabel large tissue samples for volume imaging*. Cell, 159(4): 896-910. <https://doi.org/10.1016/j.cell.2014.10.010>.
- [38] Renier, N., Adams, E., Kirst, C., Wu, Z., Azevedo, R., Kohl, J., Autry, A., Kadiri, L., Umadevi Venkataraju, K., Zhou, Y., Wang, V., Tang, C., Olsen, O., Dulac, C., Osten, P. y Tessier-Lavigne, M. (2016). *Mapping of Brain Activity by Automated Volume Analysis of Immediate Early Genes*. Cell, 165(7): 1789-1802. <https://doi.org/10.1016/j.cell.2016.05.007>.
- [39] Salkind, N. J. (2007). *Shapiro-wilk test for normality*. Encyclopedia of measurement and statistics, 0: 884-886. Sage Publications, Inc. <https://doi.org/10.4135/9781412952644.n404>.
- [40] Splittgerber, R. (2019). *Snell: Neuroanatomía clínica*, 8 ed. Lippincott Williams & Wilkins.
- [41] Stroup, W.W., Ptukhina, M., y Garai, J. (2024). *Generalized Linear Mixed Models: Modern Concepts, Methods and Applications*, 2nd ed. Chapman and Hall/CRC.
- [42] Sullivan, E. y Bigler, E. (2015). *Neuroimaging's Role in Neuropsychology: Introduction to the Special Issue of Neuropsychology Review on Neuroimaging in Neuropsychology*. Neuropsychol Review, 25(3): 221-223. <https://doi.org/10.1007/s11065-015-9296-7>.
- [43] Yen, C., Lin, C.-L. y Chiang, M.-C. (2023). *Exploring the Frontiers of Neuroimaging: A Review of Recent Advances in Understanding Brain Functioning and Disorders*. Life, 13(7): 1472. <https://doi.org/10.3390/life13071472>.