
Modelització d'outcomes basada en pertinença a perfils latents composicionals

Geòrgia Escaramís

Sessions Douglas Altman

19/06/2025



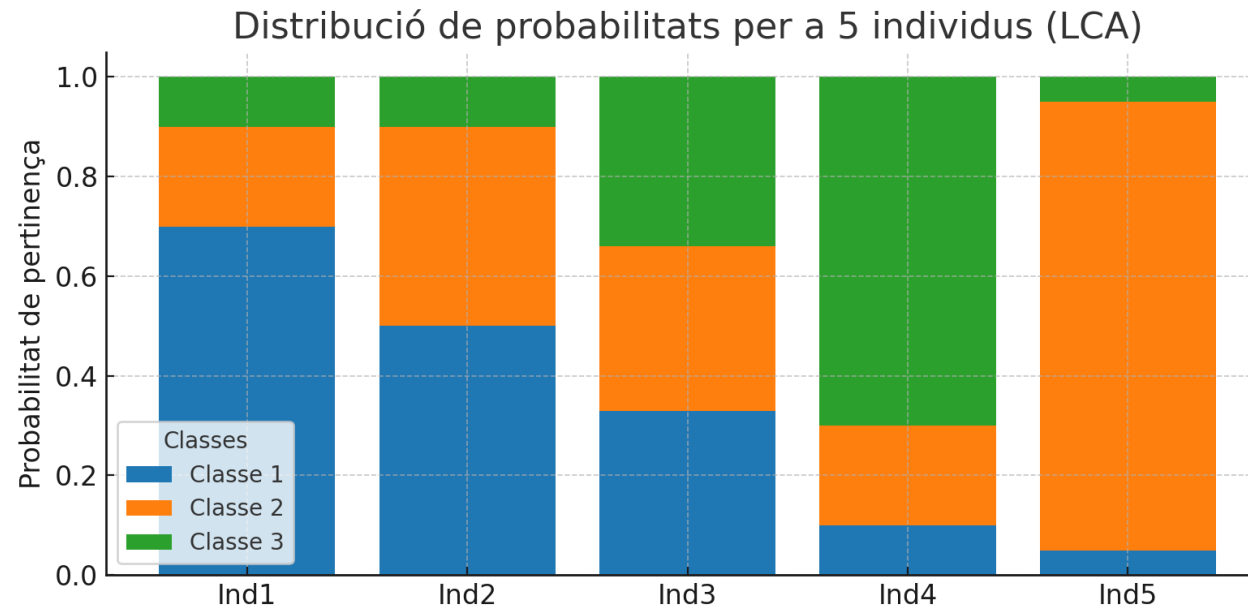
Centre d'Estudis Epidemiològics
sobre les Infeccions de Transmissió
Sexual i Sida de Catalunya

Contextualització

Per què parlem de probabilitats de pertinença?

En estudis de classes latents (LCA) o clustering probabilístic:

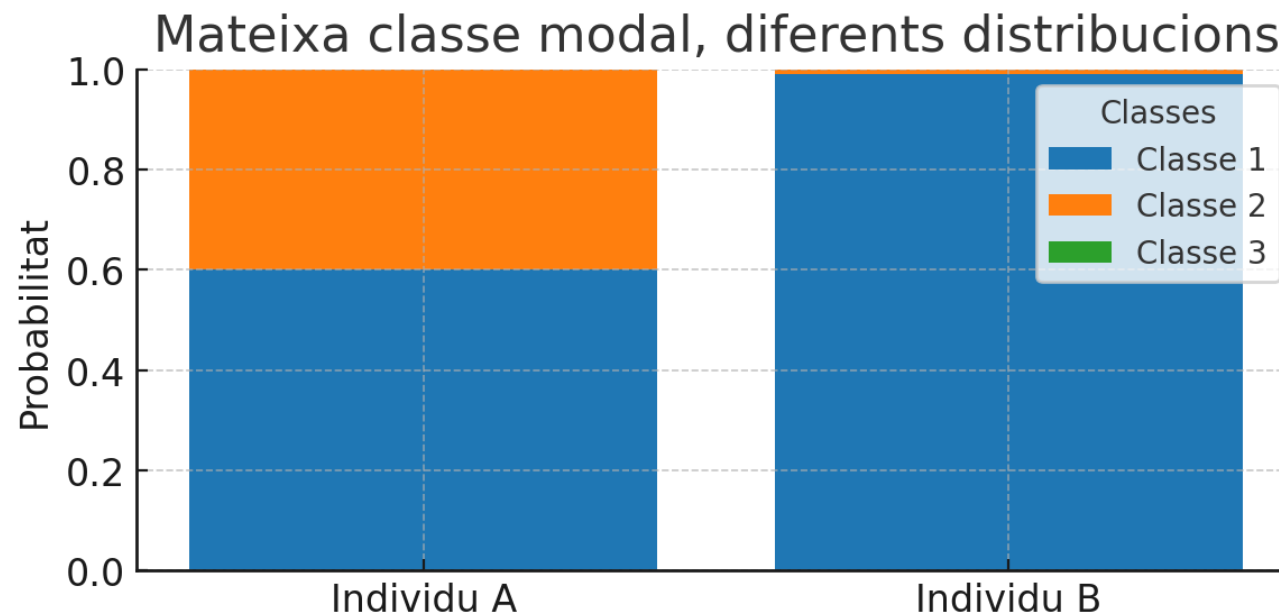
- L'objectiu és identificar perfils/segments/grups subjacents.
- El resultat no és una assignació directa, sinó una distribució de probabilitat per individu



El problema de la reducció modal

El parany del model determinista

- Assignar cada individu a la classe amb més probabilitat (modal) és temptador...
- Però això redueix un model probabilístic a un classificatori rígid.
- Estem ignorant la incertesa (entropia) i perdent informació.

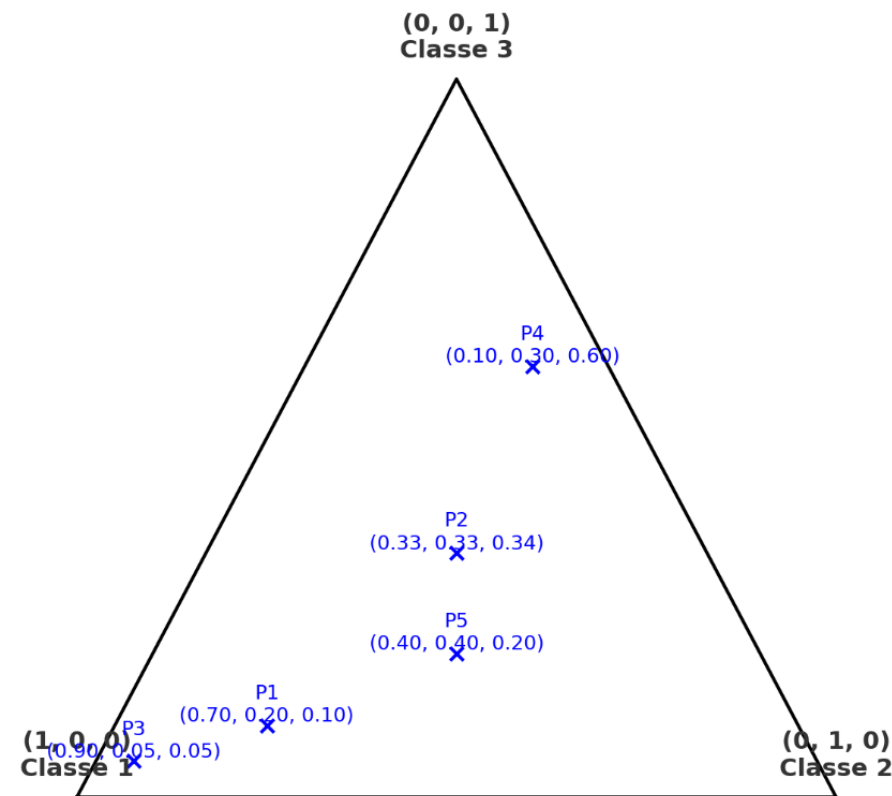


Treballar amb la composició de probabilitats

Aprofitem tota la informació

El vector de probabilitats de pertinença és una composició:

- Suma 1
- Cada component és dependent de la resta
- Està contingut dins d'un **símplex** (espai composicional)



Implicacions de la naturalesa composicional

Una variació implica compensació

- Augmentar la probabilitat d'una classe implica reduir-ne alguna altra.
- No podem treballar en termes absoluts, ni aplicar estadística clàssica directament.

Si s'ignora això... ens trobem amb:

- **Multicolinealitat** al model de regressió
- **Comparacions de mitjanes** entre grups → interpretacions “enganyoses”
- **Coeficients que semblen significatius** → artefactes composicionals
- Es perden propietats com la **interpretabilitat**

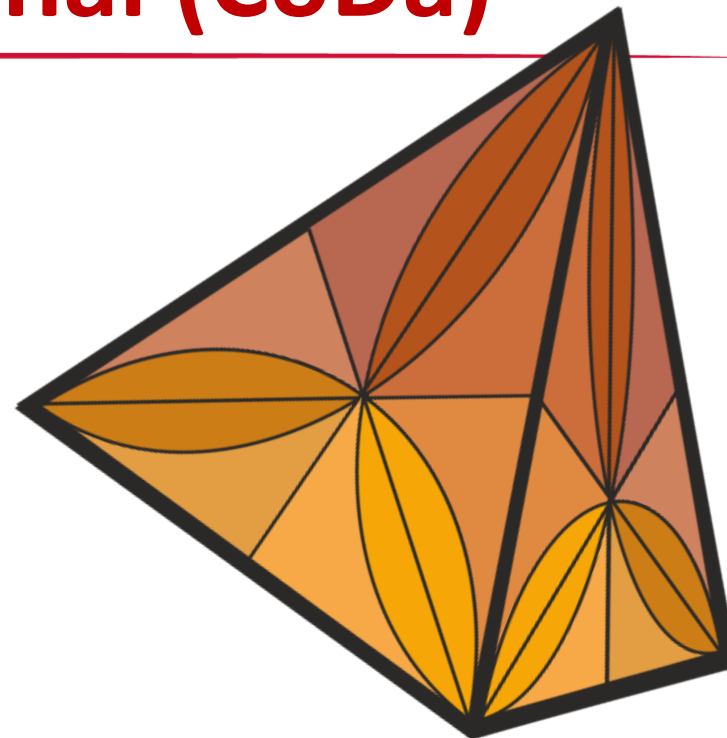
Solució: anàlisi composicional (CoDa)

Origen

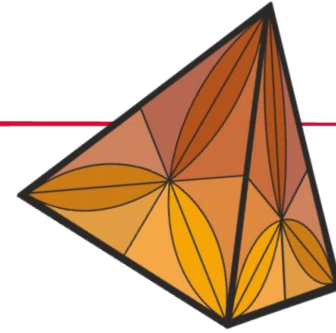
- Desenvolupada als anys 1980 per **John Aitchison**
- Inicialment aplicada a la **geologia**

Extensió a ciències de la salut:

- ✓ **Epidemiologia:** Composició del microbioma
Exposicions ambientals múltiples (metalls, contaminants)
- ✓ **Ciències clíniques:** Composició corporal (% greix, % múscul...)
Prescripcions o hàbits de consum de fàrmacs
- ✓ **Investigació social i enquestes:** Perfils latents → probabilitats de pertinença



Metodologia CoDa



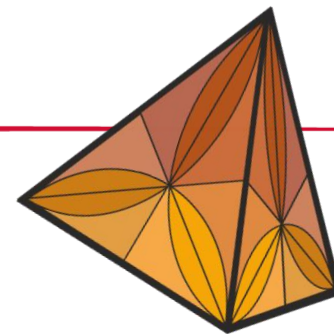
- Un vector composicional amb D components és:

$$\mathbf{x} = (x_1, x_2, \dots, x_D); x_i > 0 \text{ i } \sum_{i=1}^D x_i = k$$

exemple: $k = 1$ -> proporcions o probabilitats

- Aquest vector no viu en $\mathbb{R}^D$, sinó dins el símplex S^D
- **Objectiu:** transformació del simplex a l'espai euclidià
- **Idees clau:**
 1. Treball amb **relacions/contrastos entre parts**
 2. La **mitjana geomètrica** és la base natural: $g(\mathbf{x}) = \left(\prod_{i=1}^D x_i\right)^{1/D}$

Metodologia CoDa



Transformacions comuns

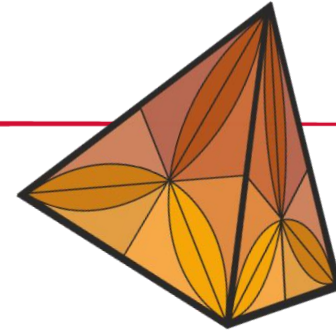
1. $clr(\mathbf{x})$ -> Log-ratio centrada

- $clr(\mathbf{x}) = \left(\ln \frac{x_1}{g(\mathbf{x})}, \dots, \ln \frac{x_D}{g(\mathbf{x})} \right), \quad g(\mathbf{x}) = \left(\prod_{i=1}^D x_i \right)^{1/D}$

2. $ilr(\mathbf{x})$ -> **Log-ratio isomètrica** -> Transformació ortonormal (amb base $\mathbf{V}$)

- $ilr(\mathbf{x}) = \mathbf{V}^T \cdot \ln(\mathbf{x})$
- La base $\mathbf{V}$ es construeix mitjançant una partició binària seqüencial (SBP)
- SBP: Cada coordenada contrasta dues agrupacions de parts (+ i -)

Metodologia CoDa



Algoritme SBP

- Grup positiu $\mathbb{R}^+$ amb r components
- Grup negatiu $\mathbb{R}^-$ amb s components
- La coordenada *ilr* és: $z = \sqrt{\frac{r \cdot s}{r+s}} \cdot \ln \left(\frac{g(x_{\mathbb{R}^+})}{g(x_{\mathbb{R}^-})} \right)$

- El vector base $v \in \mathbb{R}^D$ és: $v_i = \begin{cases} +\sqrt{\frac{s}{r(s+r)}}, & \text{si } i \in \mathbb{R}^+ \\ -\sqrt{\frac{r}{s(r+s)}}, & \text{si } i \in \mathbb{R}^- \\ 0, & \text{altrament} \end{cases}$

Exemple 3 parts

Coordenada ilr	Grup (+)	Grup (-)	Fórmula
z_1	x_1	x_2, x_3	$\sqrt{\frac{2}{3}} \cdot \ln \left(\frac{x_1}{\sqrt{x_2 x_3}} \right)$
z_2	x_2	x_3	$\sqrt{\frac{1}{2}} \cdot \ln \left(\frac{x_2}{x_3} \right)$

Aplicació en un estudi sobre mpox i estigma

Antecedents

- Les altes taxes de mpox entre homes gais, bisexuals i altres HSH (GBHSH) els exposen a més vulnerabilitat a l'estigma, fet que pot afectar la salut i dificultar la resposta de salut pública.

Objectiu

- Aquest estudi analitza com els perfils de percepció i intencionalitat envers la mpox s'associen amb creences d'estigma

Aplicació en un estudi sobre mpox i estigma

Disseny

- Estudi transversal basat en l'enquesta LIBERO-POX
- Període: administrada durant el darrer trimestre del 2022 i el primer trimestre del 2023.
- Població: homes gais, bisexuals i altres HSH (GBHSH) adults residents a Iberoamèrica.
- Mida de la mostra: $n = 8.972$

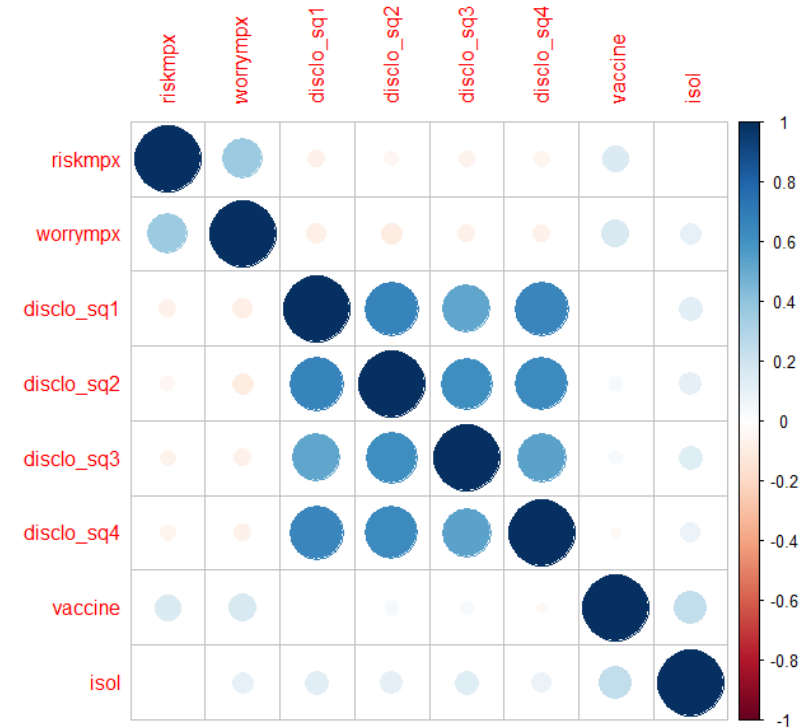
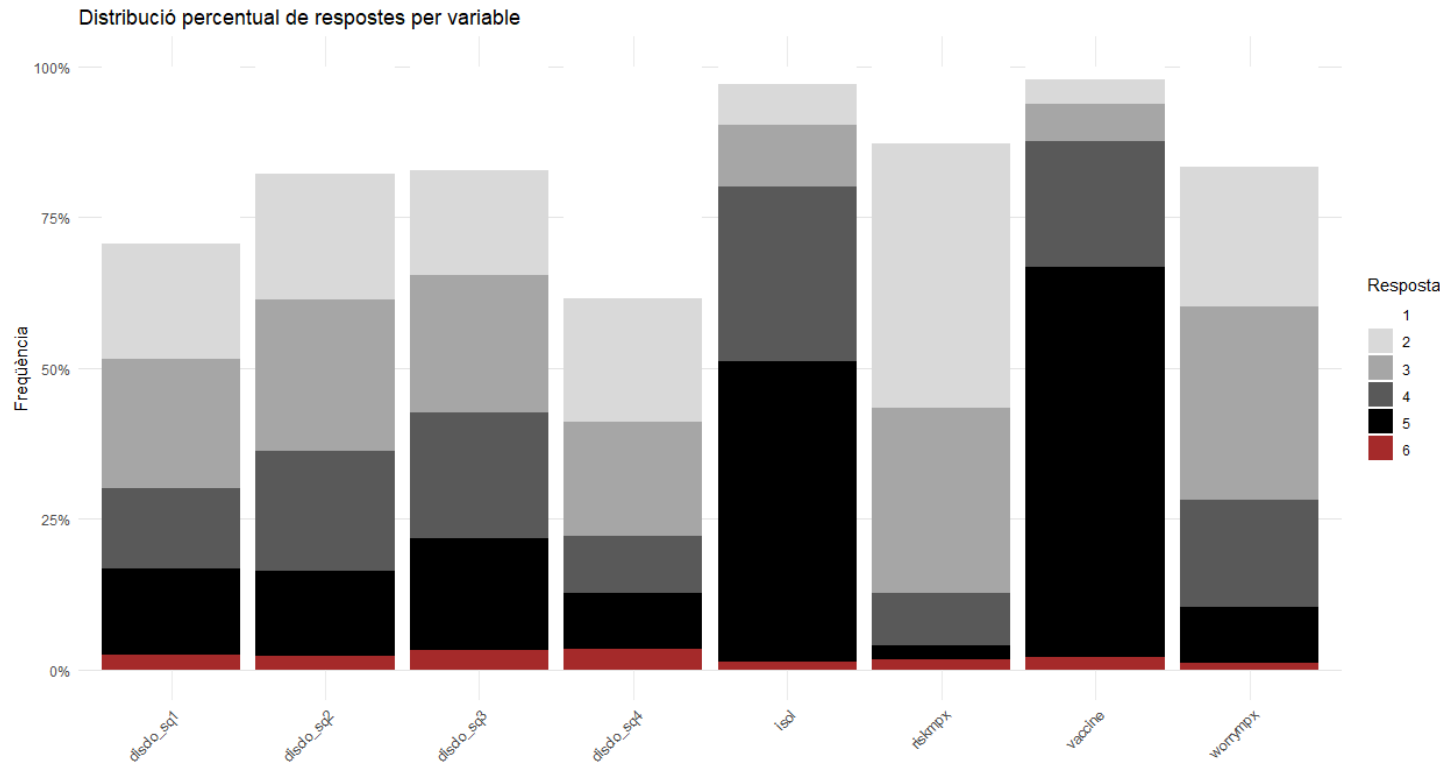
Aplicació en un estudi sobre mpox i estigma

Variables

Tipus de variable	Descripció (nom)	Escala de mesura
Variables resposta (estigma)	Estigma social percebut (snmpx)	1 = gens problemàtic – 5 = extremadament problemàtic 6 = Prefereixo no contestar
	Responsabilitat des del punt de vista propi (ownfault)	1 = gens probable – 5 = extremadament probable 6 = Prefereixo no contestar
	Responsabilitat des del punt de vista social (gaydis)	1 = gens probable – 5 = extremadament probable 6 = Prefereixo no contestar
Variables de perfil (percepció i intencionalitat)	Probabilitat percebuda d'infecció (risk)	1 = gens probable – 5 = extremadament probable 6 = Prefereixo no contestar
	Nivell de preocupació per infectar-se (worry)	1 = gens preocupat – 5 = extremadament preocupat 6 = Prefereixo no contestar
	Facilitat percebuda per comunicar el diagnòstic (disclo)	1 = gens fàcil – 5 = molt fàcil 6 = Prefereixo no contestar
	Intenció de vacunar-se (vaccine)	1 = gens probable – 5 = extremadament probable 6 = Prefereixo no contestar
	Intenció d'aïllar-se (isol)	1 = gens probable – 5 = extremadament probable 6 = Prefereixo no contestar
Variables socio-demogràfiques	Edat, país residència, educació, situació laboral	

Aplicació en un estudi sobre mpox i estigma

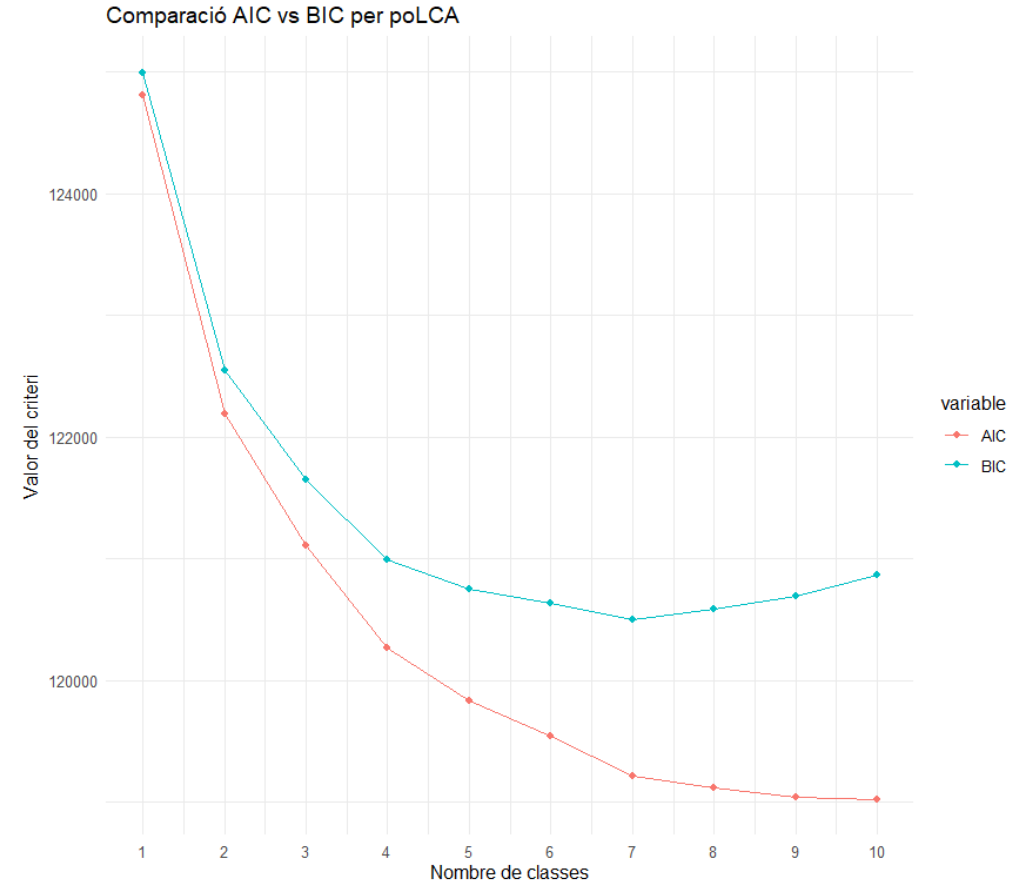
Anàlisi de perfils: pre-processament



Aplicació en un estudi sobre mpox i estigma

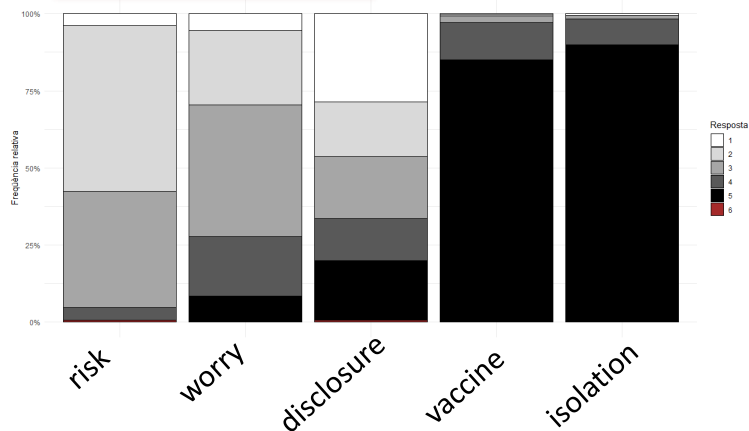
Anàlisi de perfils

- Mètode: Anàlisi de Classes Latents amb variables categòriques politòmiques
- Determinació de nombre òtim de classes: criteris AIC i BIC
- Paquet usat: poLCA de R

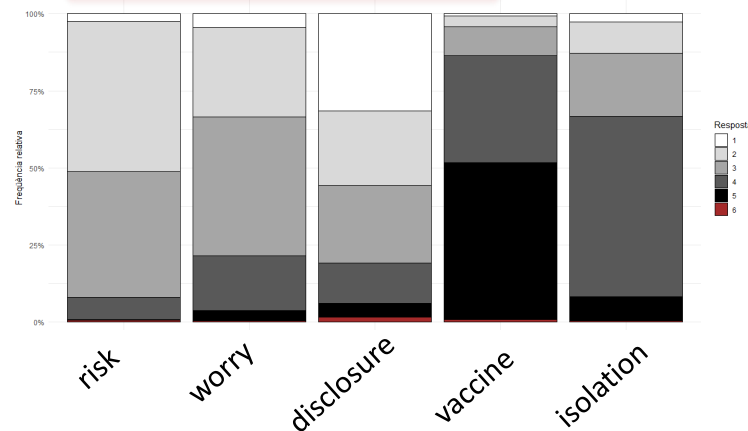


Aplicació en un estudi sobre mpox i estigma

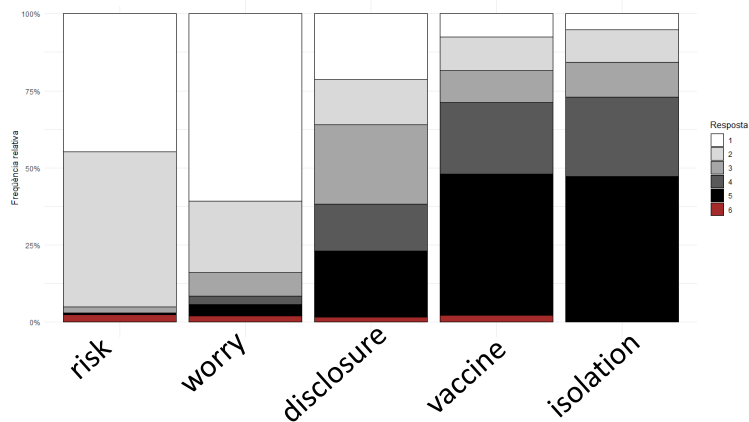
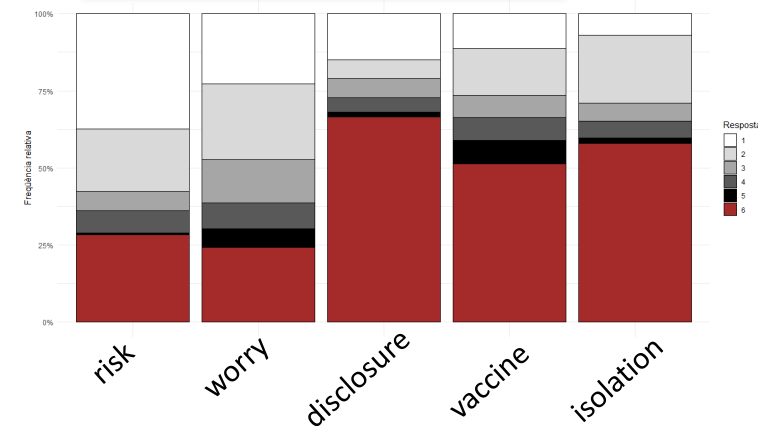
LCA1: Màxima intencionalitat



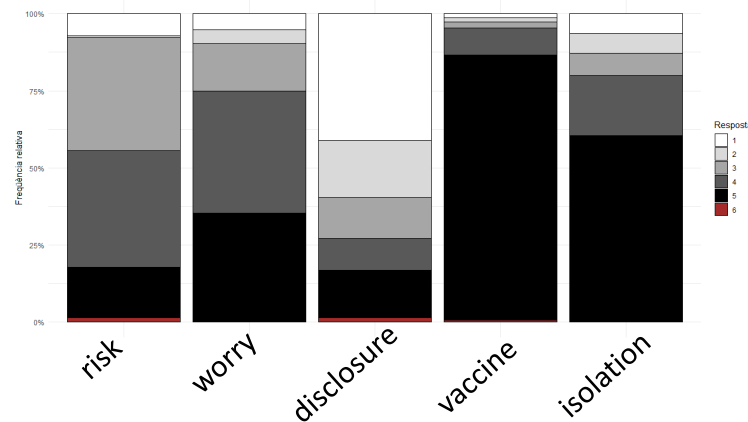
LCA2: Menor intenció d'aïllament



LCA3: No resposta o mínima intencionalitat



LCA4: Mínima preocupació



LCA5: Major preocupació, difícil comunicar-ho, bona intencionalitat

Aplicació en un estudi sobre mpox i estigma

LCA1: Màxima intencionalitat

LCA2: Menor intenció d'aïllament

LCA3: No resposta o mínima intencionalitat

LCA4: Mínima preocupació

LCA5: Major preocupació, difícil comunicar-ho, bona intencionalitat

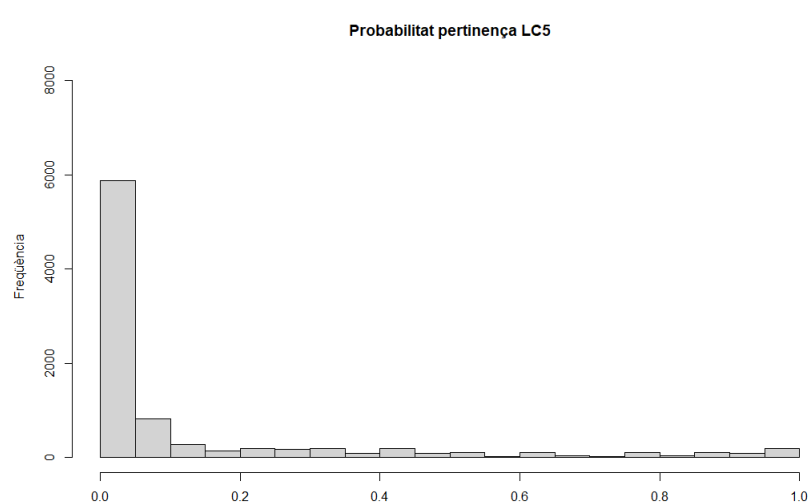
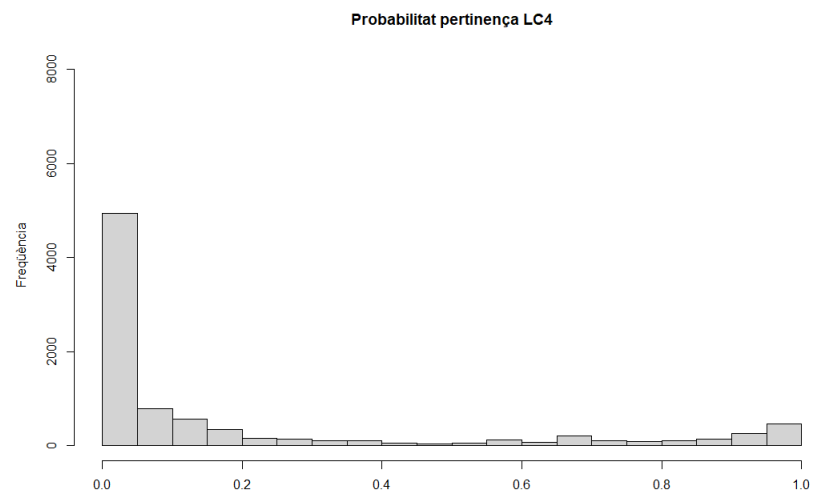
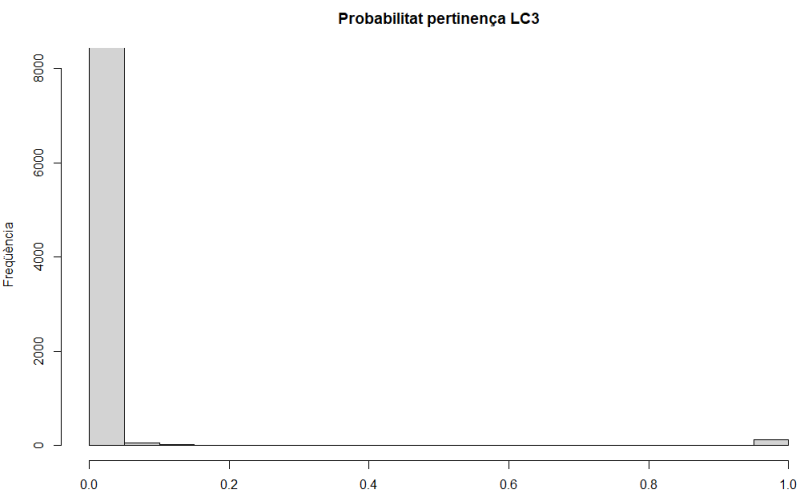
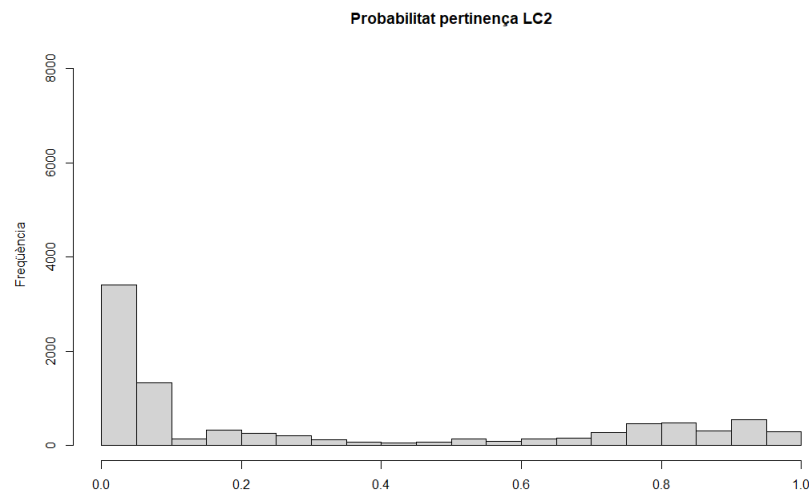
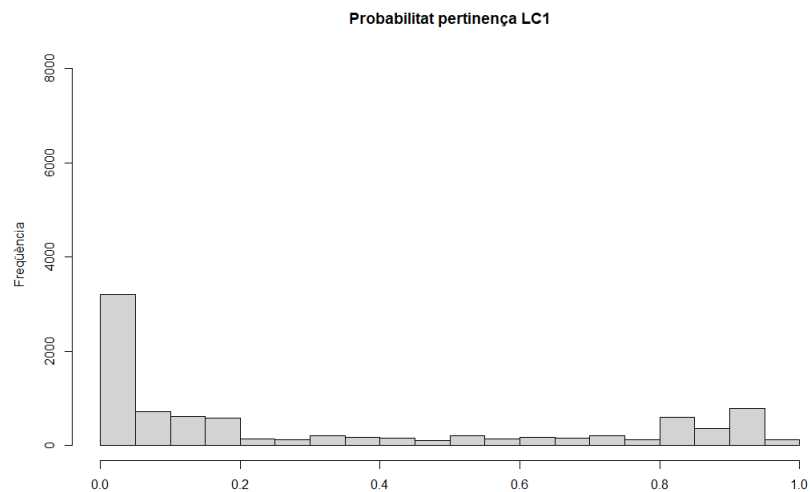
PAÍS	LCA1	LCA2	LCA3	LCA4	LCA5	Total (N)
Argentina	0.316	0.314	0.013	0.255	0.103	1098
Bolivia	0.208	0.358	0.038	0.301	0.095	176
Brasil	0.401	0.270	0.014	0.152	0.162	2689
Chile	0.353	0.302	0.009	0.237	0.099	519
Colòmbia	0.308	0.346	0.024	0.180	0.143	1228
Ecuador	0.225	0.342	0.079	0.242	0.111	886
Espanya	0.320	0.323	0.005	0.278	0.074	665
Guatemala	0.318	0.313	0.012	0.190	0.166	237
Perú	0.269	0.346	0.031	0.216	0.138	1047
Portugal	0.340	0.368	0.004	0.162	0.126	427
Total	0.327	0.315	0.023	0.203	0.131	8972

EDUCACIÓN	LCA1	LCA2	LCA3	LCA4	LCA5	Total (N)
Sin estudios o educación básica incompleta	0.138	0.340	0.043	0.280	0.198	34
Educación básica o estudios primarios	0.140	0.327	0.112	0.304	0.117	91
Estudios secundarios	0.275	0.315	0.041	0.249	0.119	1223
Estudios de bachillerato o FP	0.298	0.336	0.024	0.235	0.107	1856
Estudios universitarios	0.354	0.307	0.014	0.181	0.144	4138
Máster o doctorado	0.358	0.318	0.006	0.180	0.138	1528
Prefiero no contestar	0.114	0.190	0.395	0.211	0.089	69
Total	0.327	0.315	0.023	0.203	0.131	8939

SITUACIÓN LABORAL	LCA1	LCA2	LCA3	LCA4	LCA5	Total (N)
Empleado, tiempo completo	0.345	0.317	0.014	0.184	0.140	4608
Empleado, medio tiempo	0.280	0.339	0.036	0.234	0.110	625
Trabajador independiente/autónomo	0.316	0.330	0.022	0.198	0.134	1593
Estudiante	0.309	0.317	0.027	0.249	0.098	901
Jubilado	0.313	0.250	0.041	0.289	0.107	135
Desempleado	0.332	0.288	0.016	0.228	0.136	793
Recibo asistencia social	0.313	0.195	0.026	0.261	0.205	45
Total	0.327	0.315	0.023	0.203	0.131	8700

EDAD	LCA1	LCA2	LCA3	LCA4	LCA5	Total (N)
18-24	0.285	0.332	0.038	0.248	0.098	1435
25-29	0.336	0.305	0.026	0.210	0.123	1647
30-34	0.328	0.329	0.022	0.180	0.141	1666
35-39	0.338	0.300	0.016	0.188	0.158	1372
40-44	0.339	0.312	0.011	0.180	0.158	972
45-49	0.345	0.301	0.010	0.206	0.138	651
50-59	0.344	0.325	0.014	0.202	0.116	745
60+	0.351	0.291	0.023	0.228	0.107	183
Total	0.327	0.315	0.023	0.203	0.131	8671

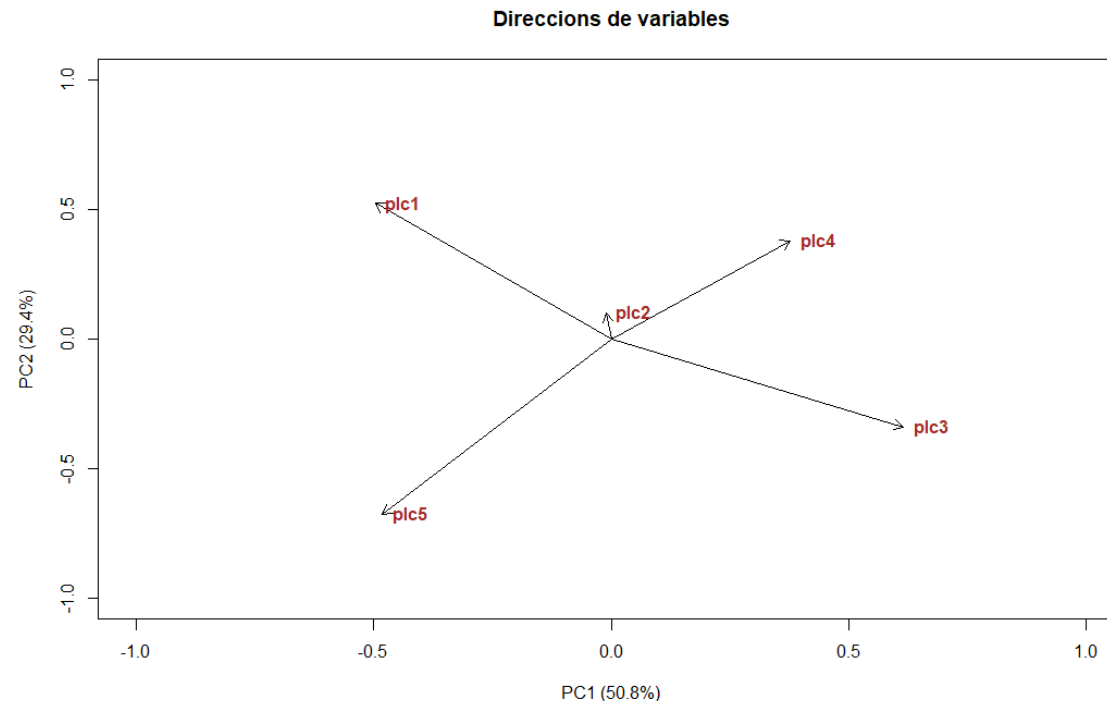
Aplicació en un estudi sobre mpox i estigma



Aplicació en un estudi sobre mpox i estigma

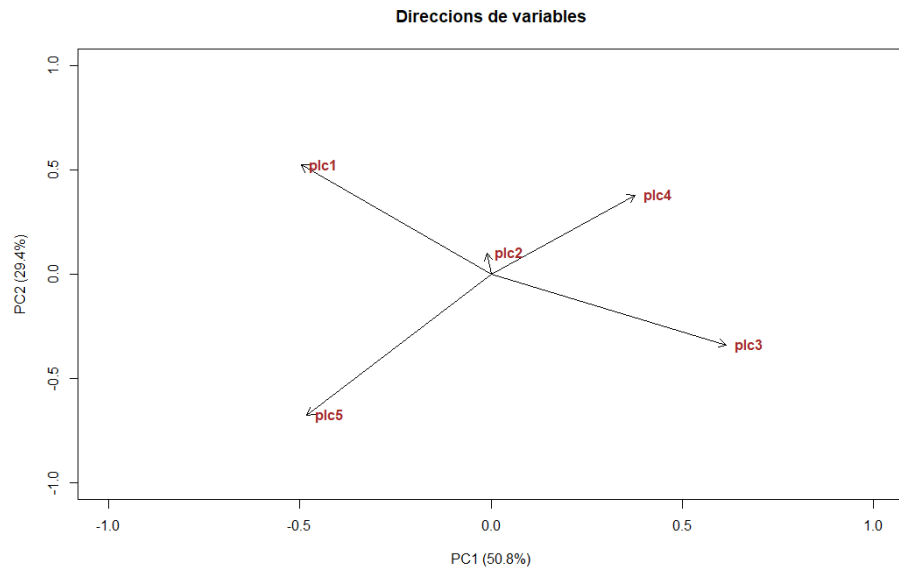
Anàlisi composicional prèvia al modelatge

- Objectiu: Seleccionar coordenades isomètriques rellevants
- Transformació de la pertinença (pLC) a CLR (log-ratios centrades)
- Anàlisi de components principals (per avaluar estructures internes)



Aplicació en un estudi sobre mpox i estigma

Construcció de les coordenades isomètriques: llibreires zComposition i compositions d'R



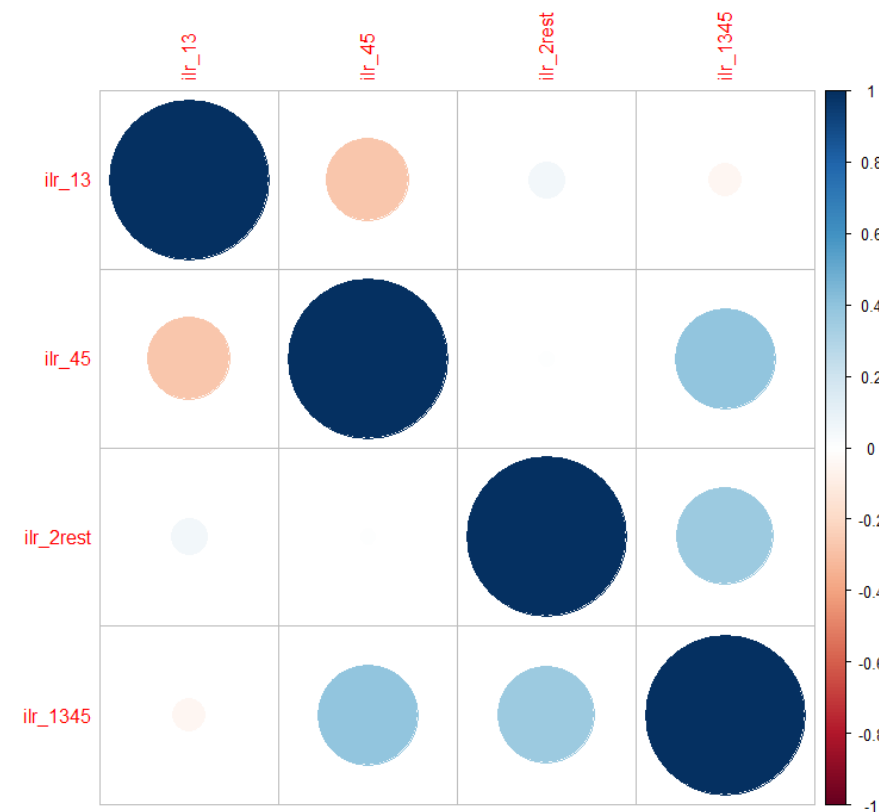
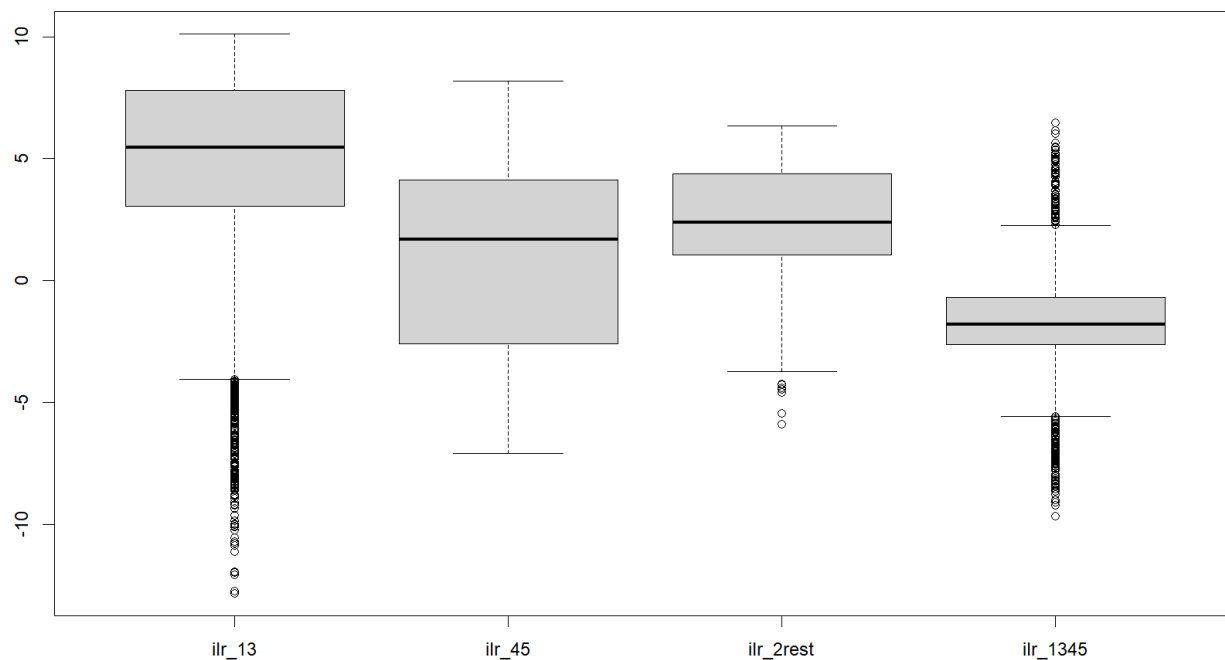
Coordenada ilr	Grup (+)	Grup (-)	Fórmula
ilr_{2_rest}	pcl_2	$pcl_1, pcl_3, pcl_4, pcl_5$	$\sqrt{\frac{4}{5}} \ln \left(\frac{pcl_2}{\sqrt[4]{pcl_1 \cdot pcl_3 \cdot pcl_4 \cdot pcl_5}} \right)$
ilr_{13_45}	pcl_1, pcl_3	pcl_4, pcl_5	$\ln \left(\frac{\sqrt{pcl_1 \cdot pcl_3}}{\sqrt{pcl_4 \cdot pcl_5}} \right)$
ilr_{1_3}	pcl_1	pcl_3	$\sqrt{\frac{1}{2}} \ln \left(\frac{pcl_1}{pcl_3} \right)$
ilr_{4_5}	pcl_4	pcl_5	$\sqrt{\frac{1}{2}} \ln \left(\frac{pcl_4}{pcl_5} \right)$

Contrastos forts:

- LCA1 (màxima intencionalitat) versus LCA3 (no resposta o mínima intencionalitat)
- LCA4 (mínima preocupació) versus LCA5 (major preocupació, comunicació difícil, bona intencionalitat)

Aplicació en un estudi sobre mpox i estigma

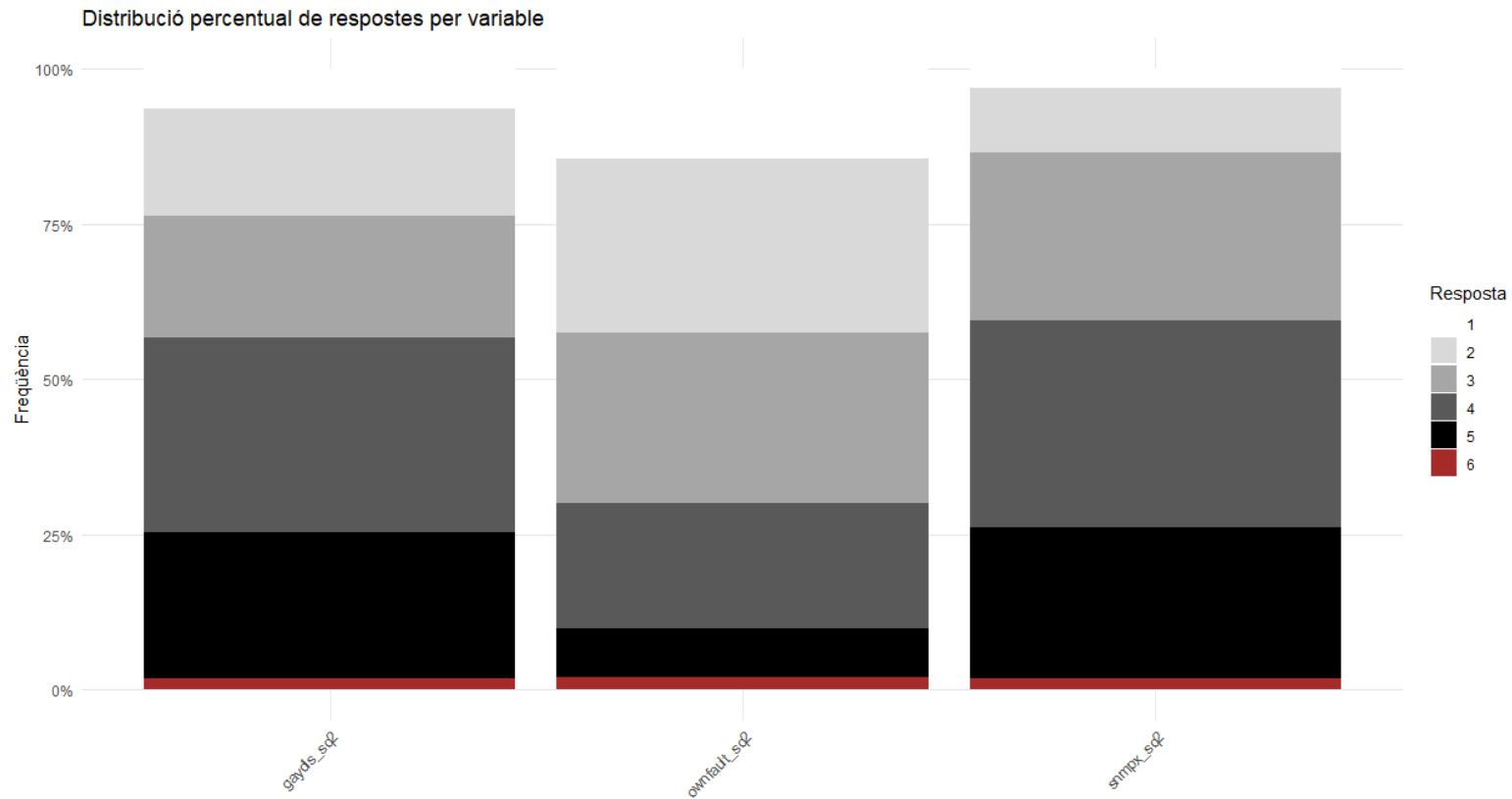
Construcció de les coordenades isomètriques



Aplicació en un estudi sobre mpox i estigma

Modelització amb coordenades ilr

- Objectiu: Efecte dels contrastos composicionals sobre l'estigma.



Aplicació en un estudi sobre mpox i estigma

Modelització amb coordenades ilr

Estratègia de modelatge en dos passos

1. Objectiu: determinar si els contrastos composicionals s'associen amb la probabilitat de no respondre.

No resposta (model logístic + matching).

Ajust pel desequilibri: propensity score matching

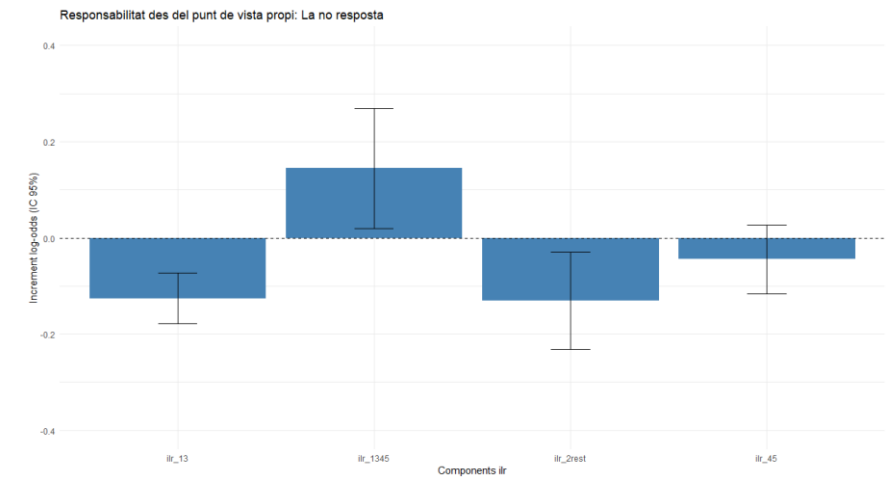
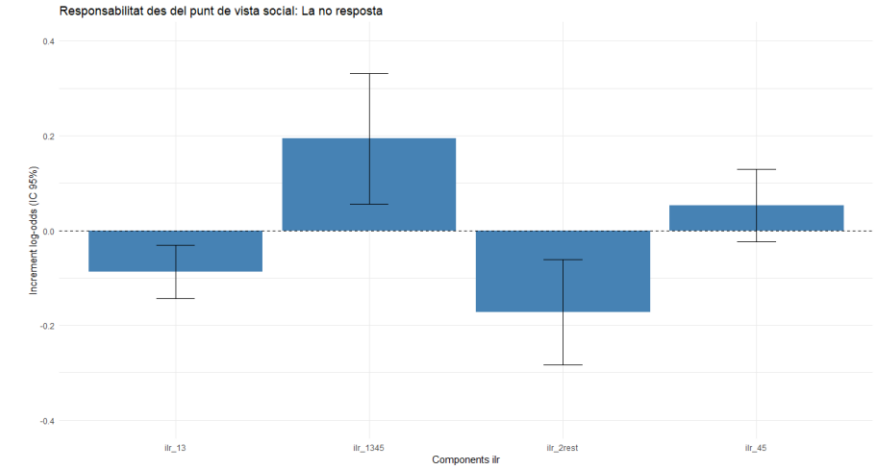
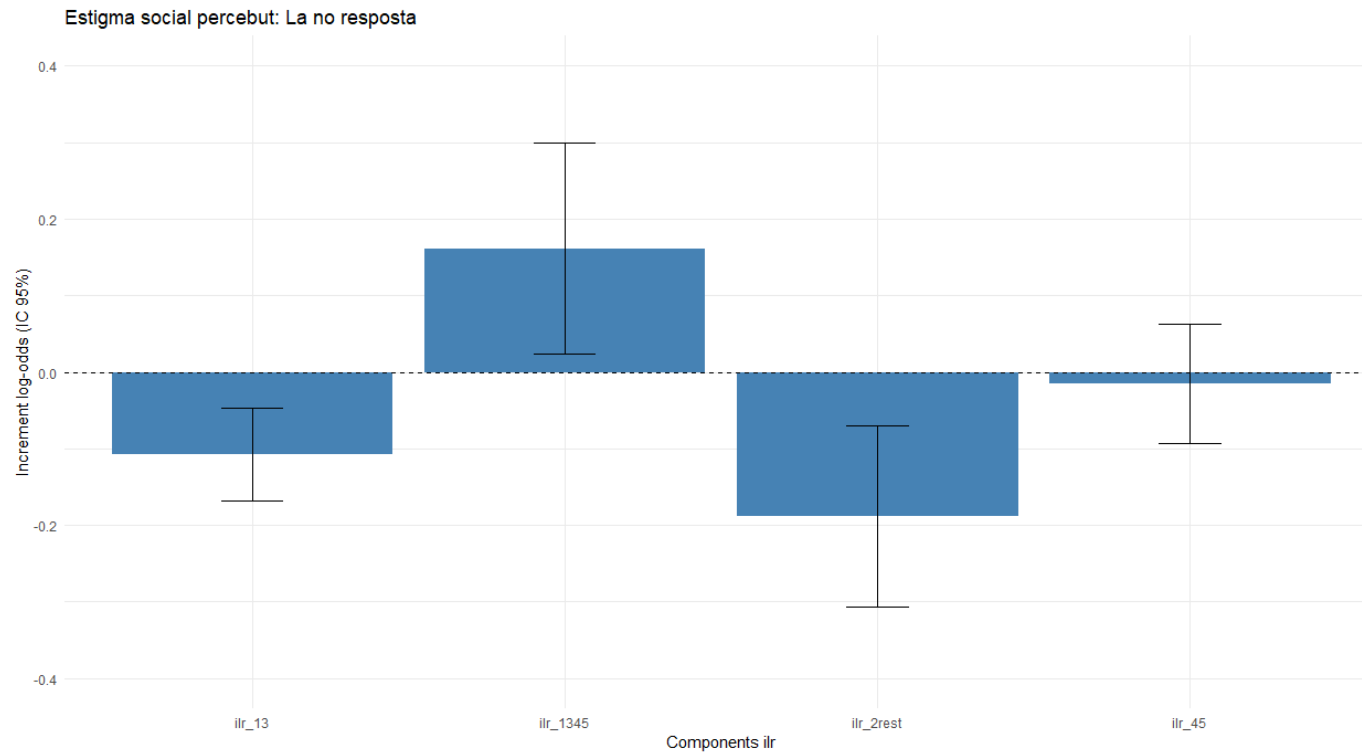
2. Objectiu: avaluar l'associació dels contrastos composicionals i el nivell d'estigma (escala Likert).

Intensitat de resposta (model lineal ajustat).

Ajust per variables sociodemogràfiques

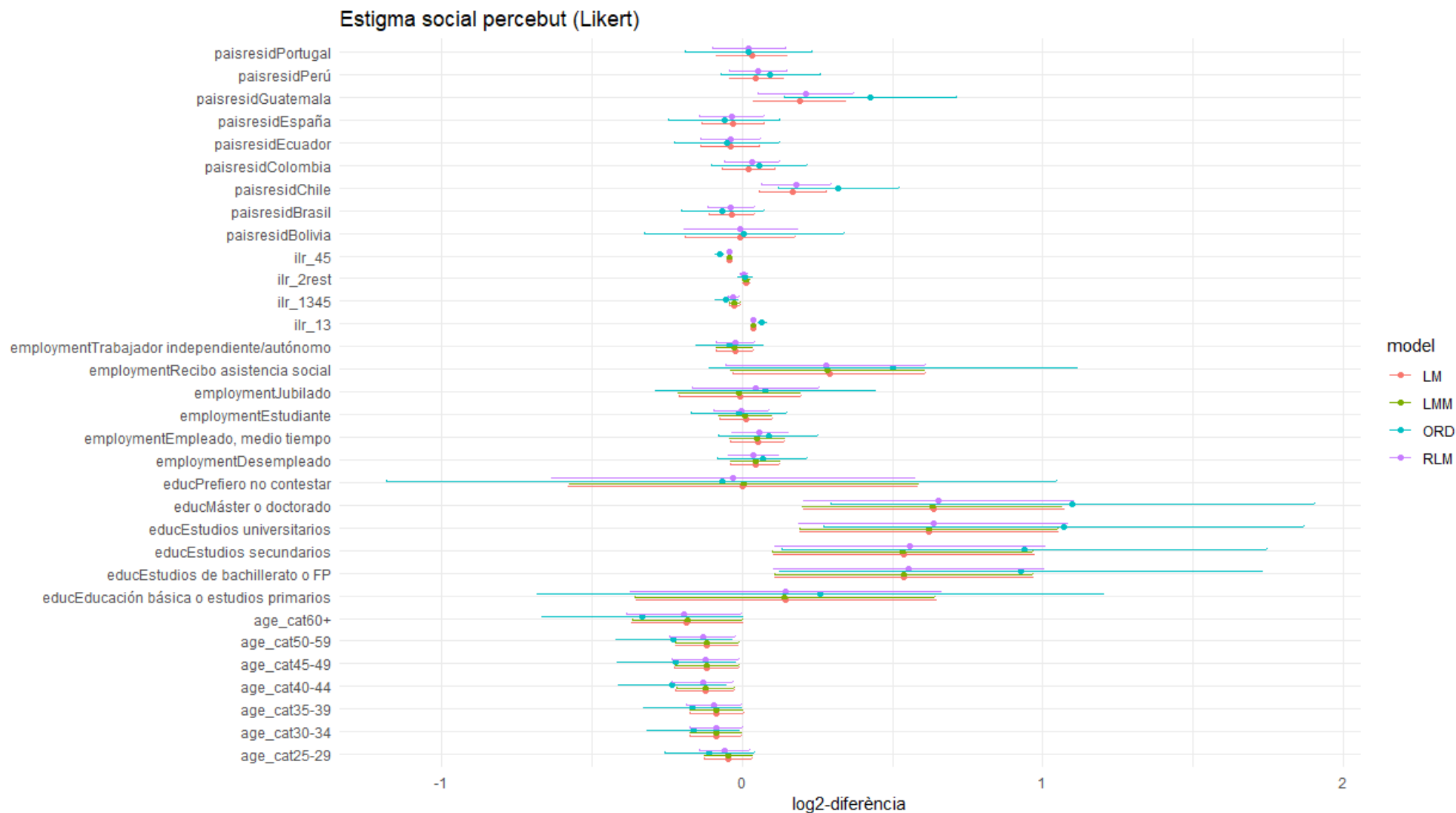
Aplicació en un estudi sobre mpox i estigma

Modelització amb coordenades ilr



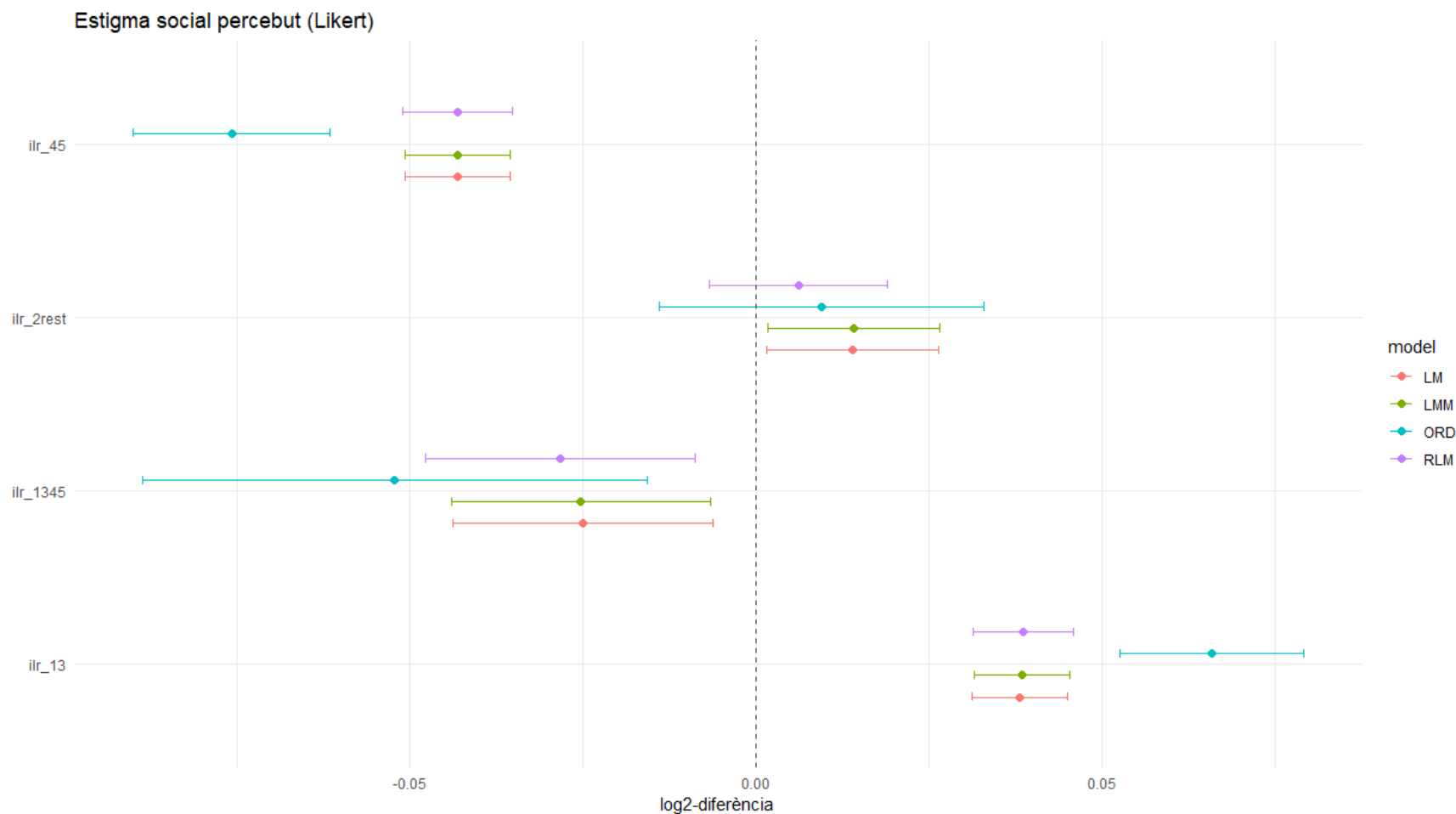
Aplicació en un estudi sobre mpox i estigma

Modelització amb coordenades ilr



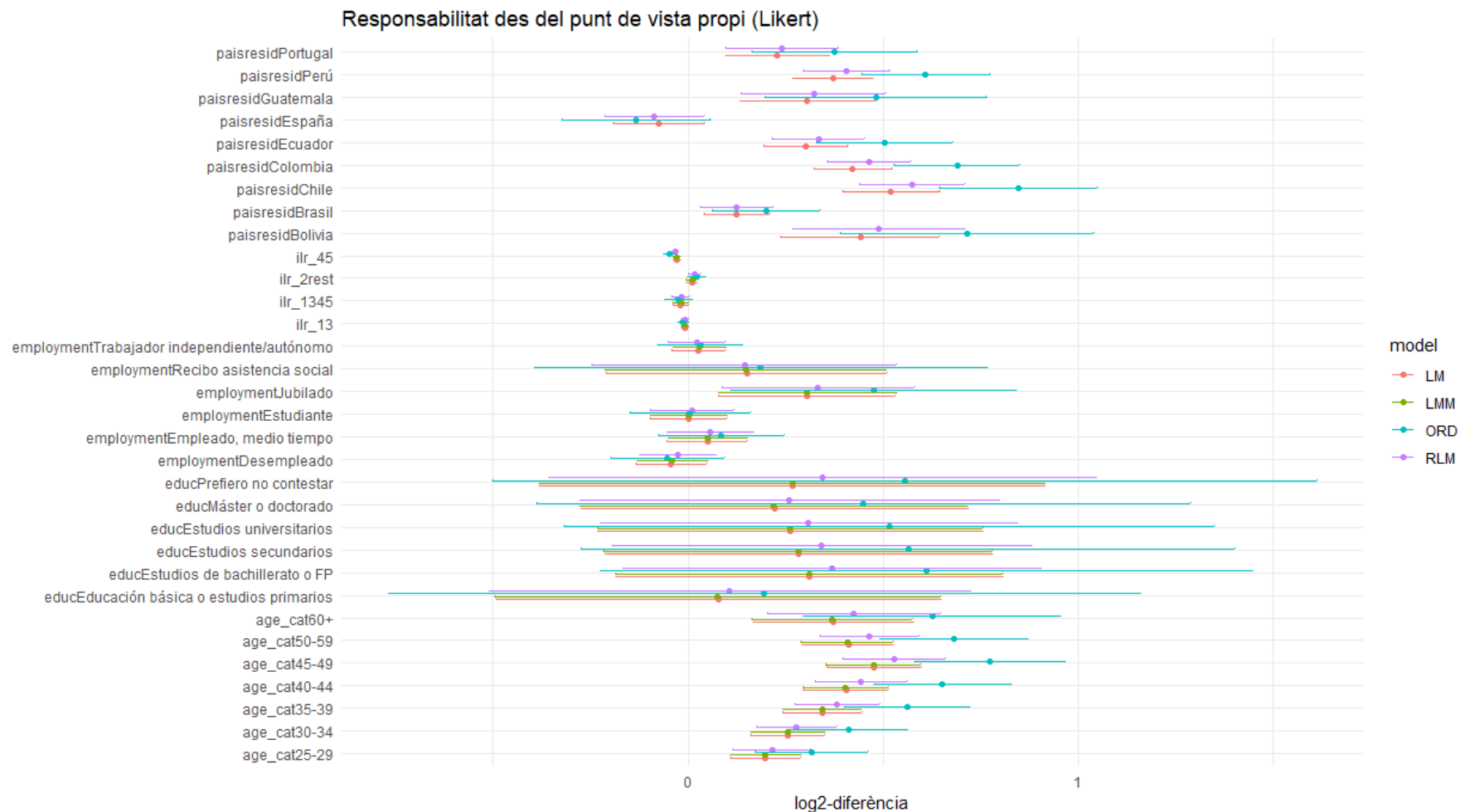
Aplicació en un estudi sobre mpox i estigma

Modelització amb coordenades ilr



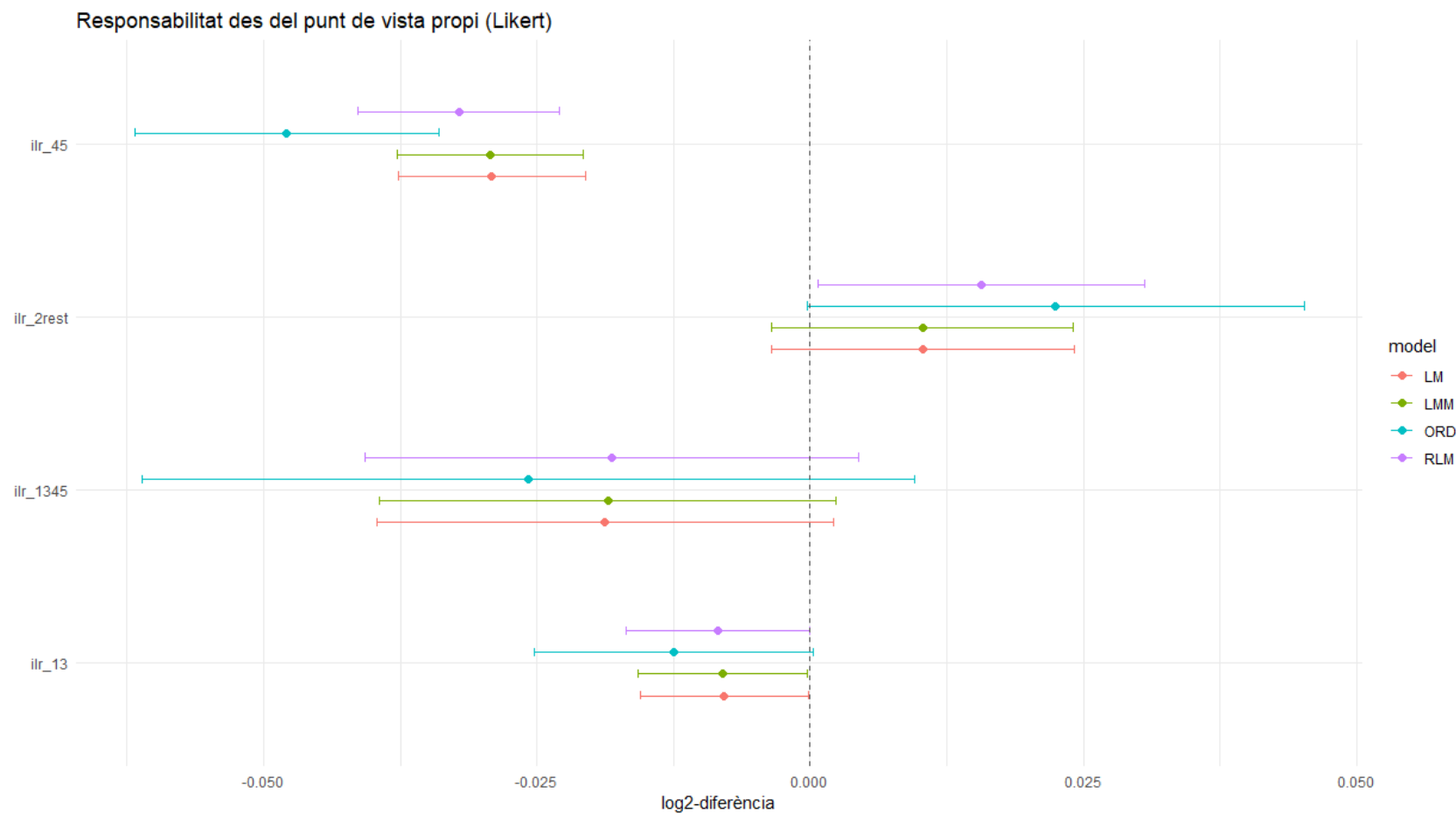
Aplicació en un estudi sobre mpox i estigma

Modelització amb coordenades ilr



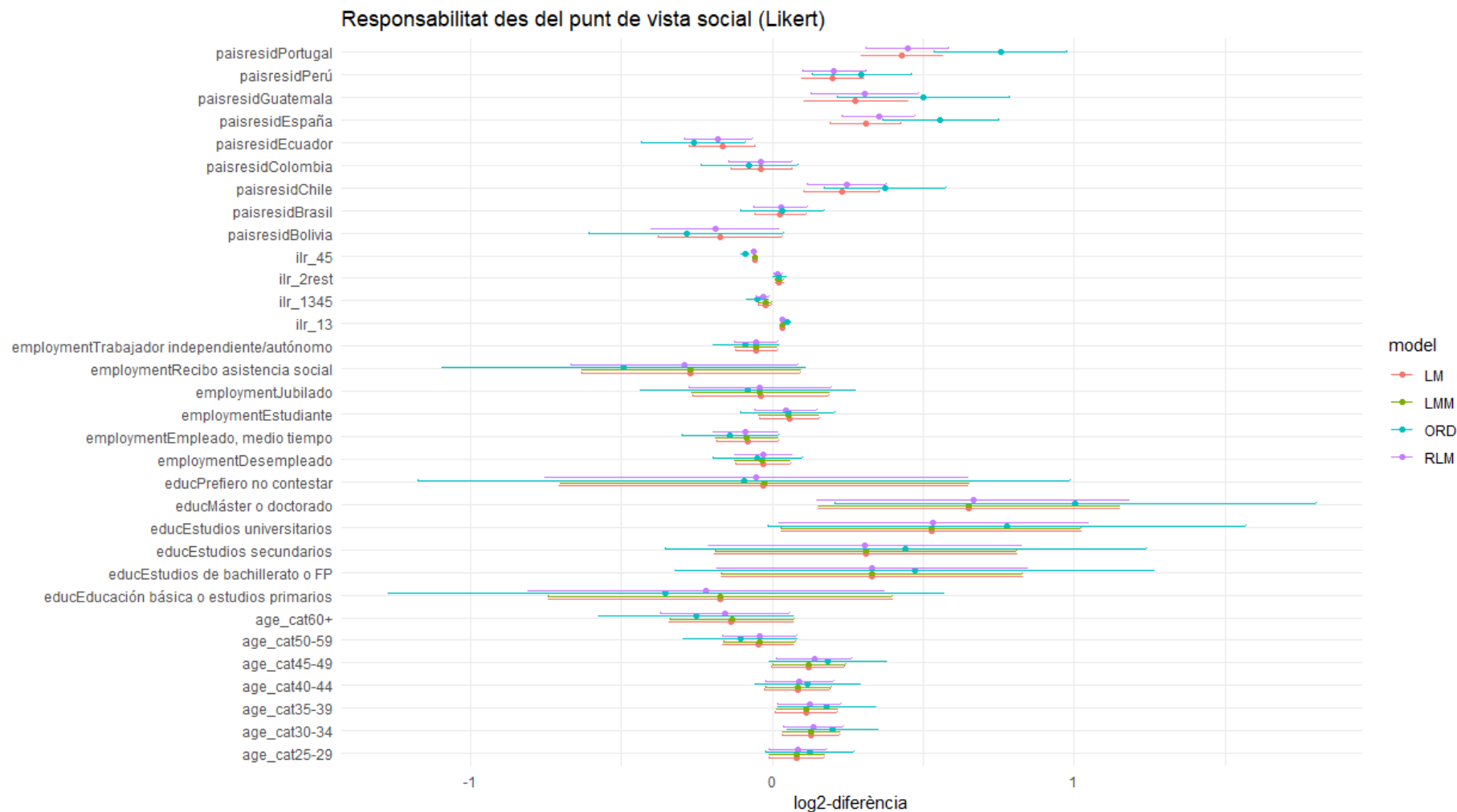
Aplicació en un estudi sobre mpox i estigma

Modelització amb coordenades ilr



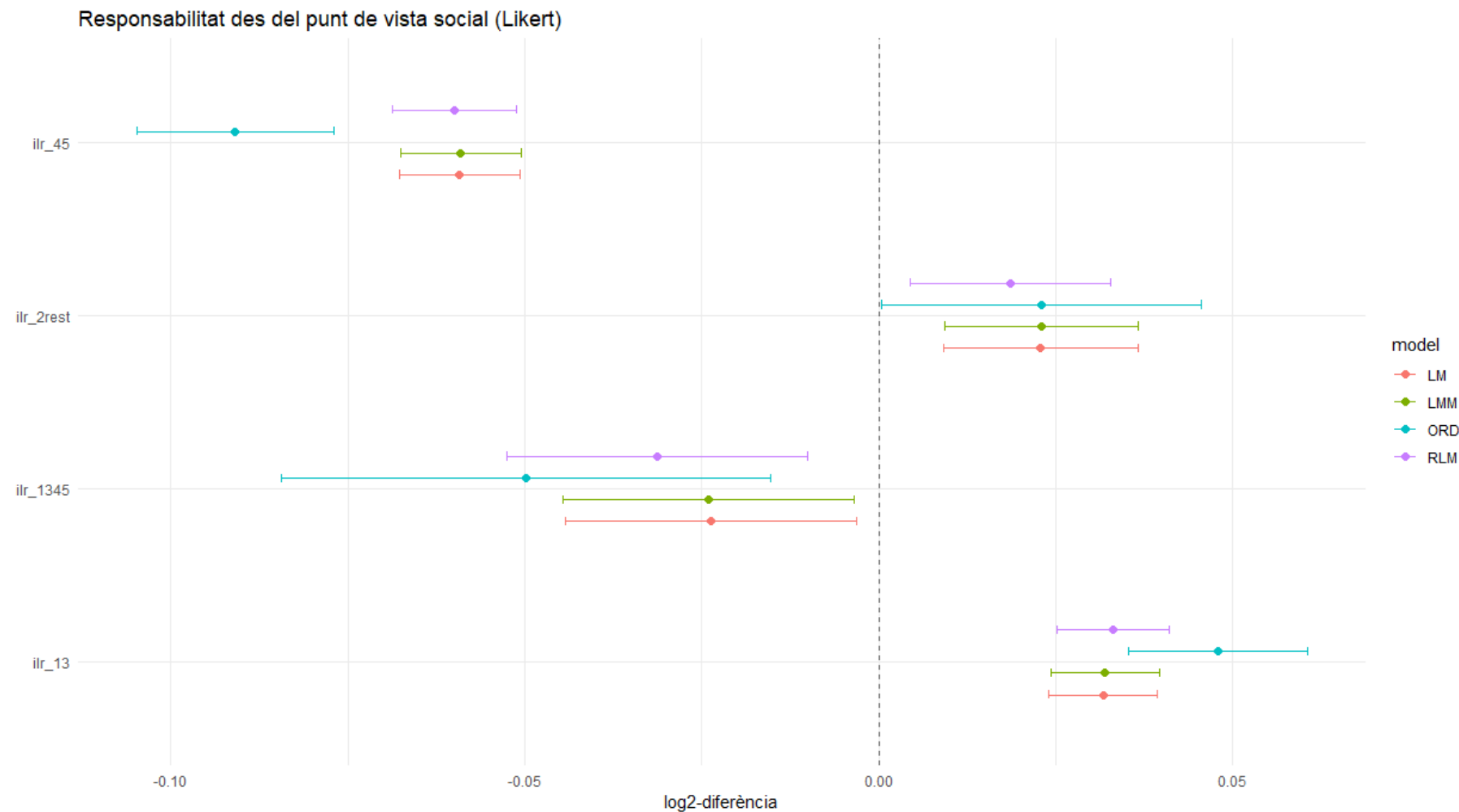
Aplicació en un estudi sobre mpox i estigma

Modelització amb coordenades ilr



Aplicació en un estudi sobre mpox i estigma

Modelització amb coordenades ilr



Gràcies!

