

1. OBJETIVOS

Aplicación de GLM y sus extensiones (Modelos lineales Generalizados Mixtos y Modelos Aditivos Generalizados) tanto en el ámbito médico como en el actuarial para predecir la probabilidad de que la mujer quede embarazada y la probabilidad de muerte respectivamente. Los modelos obtenidos se compararán mediante las curvas ROC y seleccionará el mejor modelo para cada ámbito. Además se propondrá un punto de corte en las puntuaciones que ofrece el mejor modelo para poder clasificar a los distintos individuos. Con todo ello se pretende proponer una nueva metodología en cada campo, diferente a las propuestas hasta ahora para facilitar la decisión del usuario final, a la vez que objetivarla. Todos los análisis estadísticos, se implementarán con código R utilizando muchas librerías que ponen a nuestro alcance las más novedosas herramientas estadísticas. Hemos seleccionado este software por la precisión, exactitud en sus resultados y por ser un software libre, y por tanto, de fácil implantación al alcance de todos.

2. ETAPAS PRINCIPALES DEL DESARROLLO DE LA INVESTIGACIÓN

1. Búsqueda bibliográfica y de las librerías necesarias
2. Depuración y análisis de las bases de datos
3. Desarrollo de varios modelos adecuados a la naturaleza y número de nuestras variables para cada campo
4. Comparación de modelos utilizando las curvas ROC y cálculo de sus correspondientes AUC
5. A partir de las puntuaciones que ofrece el mejor modelo, se propone un punto de corte óptimo para clasificar a los individuos en cada campo, de tal manera que esta clasificación se acerque lo máximo posible a los valores observados

3. RESULTADOS

3.1 Campo médico: Predicción de la probabilidad de embarazo utilizando técnicas de reproducción asistida

A continuación se comparan los modelos mediante las curvas ROC en ambas muestras utilizando la librería ROCC.

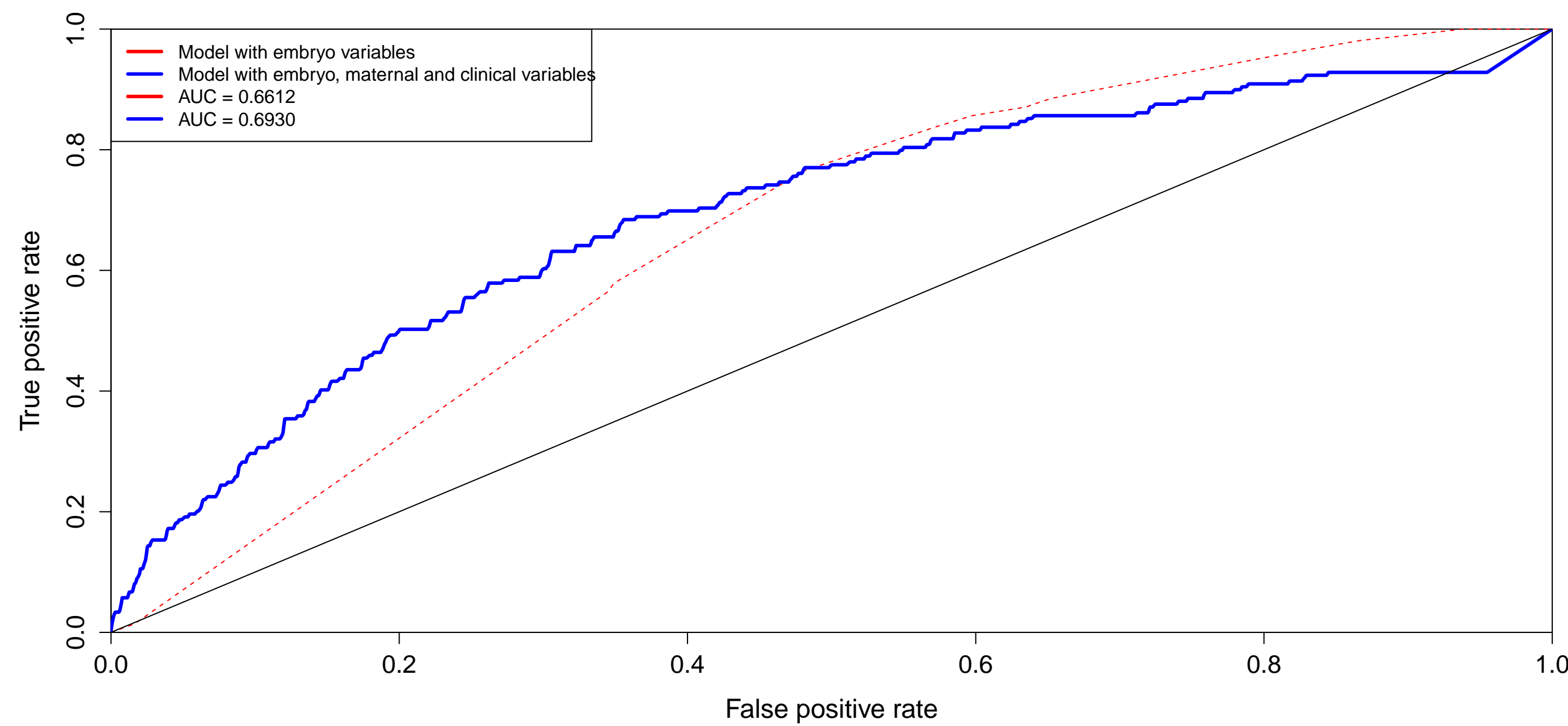


Figure 1: Comparación de curvas ROC en la muestra de construcción

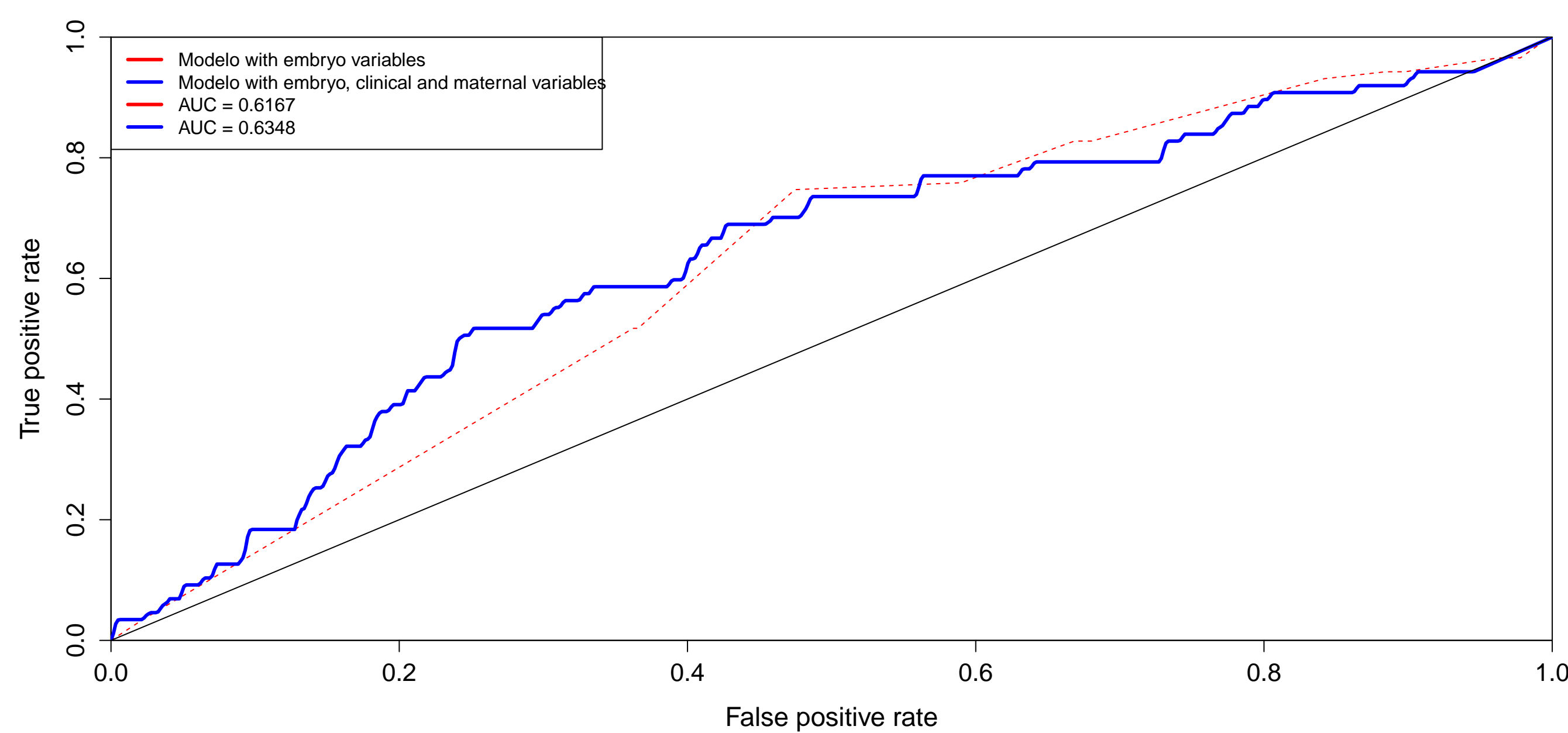


Figure 2: Comparación de curvas ROC en la muestra de validación

Comparación de las AUC de los dos modelos en ambas muestras utilizando la librería pROC.

Muestra	Método	D=(AUC1-AUC2 / σ)	p-value	AUC roc modelo1	AUC roc modelo2
Construcción	Bootstrap	-4.56	5.02E-06	66.61	73.28
Construcción	Bootstrap	-3.45	0.0005506	50.50	65.29

Se puede observar tanto en las gráficas como en la tabla anterior, que el modelo 2 (variables embrionarias, maternas y clínicas) se selecciona como mejor modelo.

Después de seleccionar el mejor modelo, se calculó el punto de corte óptimo (OOP) con el fin de distinguir entre las mujeres embarazadas y no embarazadas. Este umbral se calcula como aquel cuya distancia vertical a la diagonal fue mayor, es decir, aquel que maximice el Índice de Youden (J). Este índice está muy extendido en el campo de la medicina.

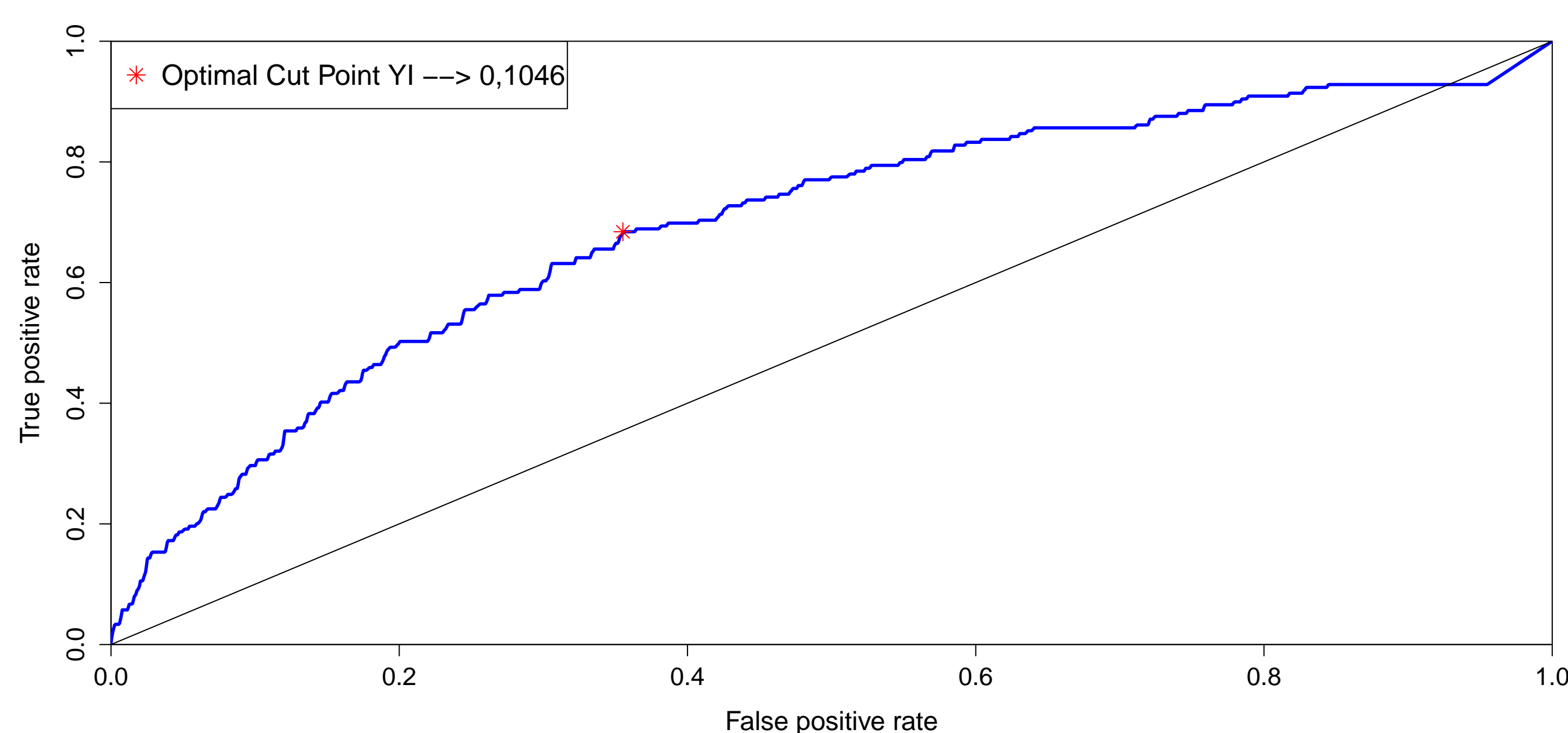


Figure 3: OOP del modelo modelo 2 (variables embrionarias, maternas y clínicas) en la muestra de construcción

En conclusión, el modelo 2 se selecciona como mejor modelo por tener mayor AUC. Destacar, que las curvas ROC se cruzan las unas sobre las otras. Un resultado positivo en la predicción de no embarazo podría implicar asesoramiento negativo en un programa de FIV. En este caso, el impacto de un resultado positivo falso es claramente más dramático. Una tasa de falsos positivos requiere una excelente especificidad, lo cual sólo puede obtenerse a expensas de una disminución de la sensibilidad. Por este motivo, volvemos a seleccionar de modelo 2 como mejor modelo. Además, hemos propuesto las curvas ROC como una herramienta gráfica para la comparación de los diferentes modelos y las AUC y J como valores numéricos resumen de las mismas. Este modelo también puede ser utilizado en aquellas bases de datos en la que, las variables igualdad y simetría se agrupan en la variable grado.

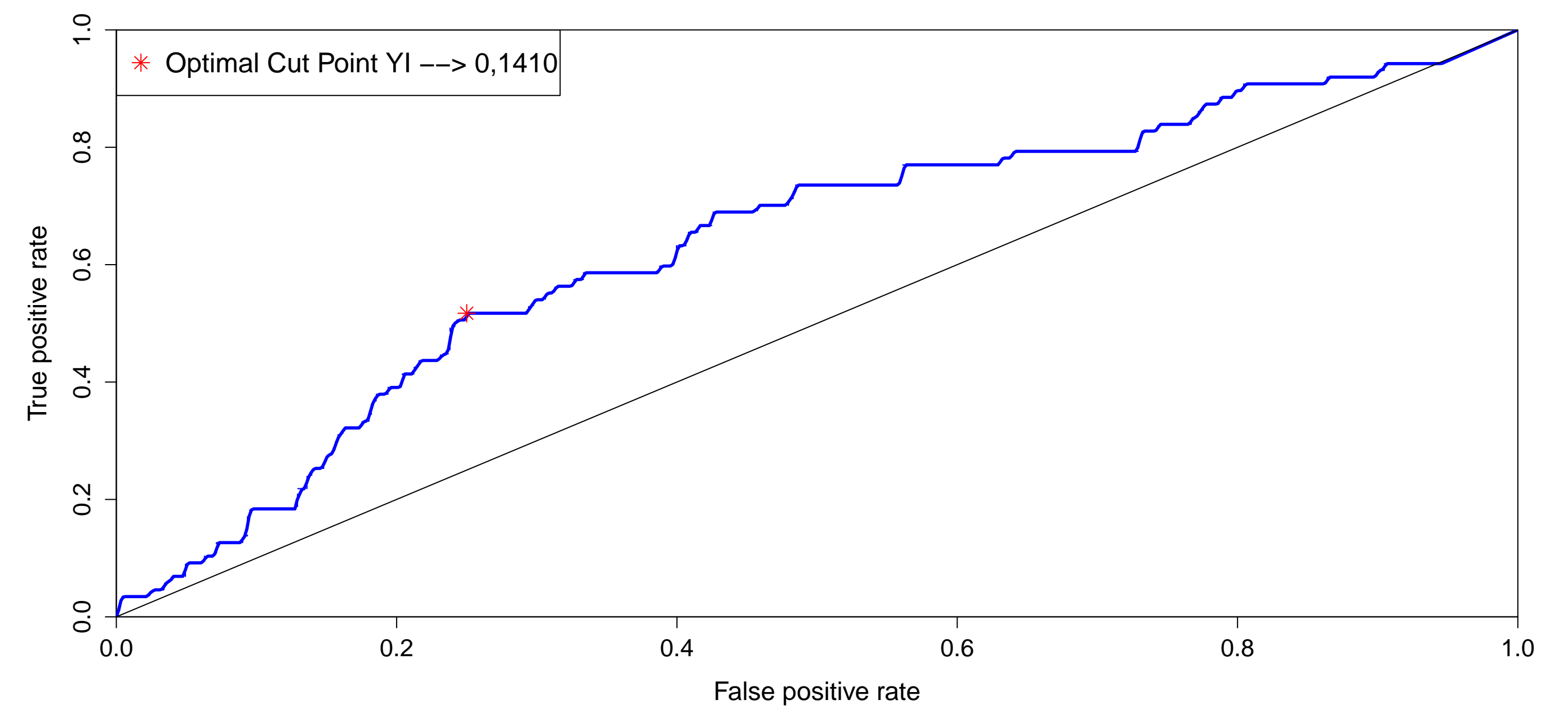


Figure 4: OOP del modelo modelo 2 (variables embrionarias, maternas y clínicas) en la muestra de construcción

3.2 Campo actuarial: Predicción de la probabilidad de muerte en países de territorio europeo por sexo.

La mortalidad ha disminuido en todos los países de la Unión Europea durante el siglo pasado, presentando patrones similares en el cambio de la misma. A pesar de estas tendencias similares, todavía hay diferencias considerables en los niveles de mortalidad, así como entre hombres y mujeres. El objetivo de este estudio es presentar un método para la detección de grupos de países en territorio de la Unión Europea con una mortalidad similar diferenciando entre sexos. El método se basa en el análisis de datos funcionales teniendo en cuenta la ubicación geográfica y, por consiguiente, las relaciones de vecindad entre ellos.

Una vez obtenidos los clusters de países crearemos un modelo común para cada uno de ellos, en vez de crear un modelo para cada uno de los países. Para ello, utilizaremos los GLM y sus extensiones, teniendo en cuenta el efecto aleatorio espacial y temporal.

La gama de colores verde o valores inferiores a 1 de la Figura 5 representa los países con un defecto de la mortalidad en comparación con sus países vecinos. Por otro lado, la gama de color marrón o valores superiores a 1 representan los países con un exceso de mortalidad. Por lo tanto, para las mujeres, Francia, Italia, España, Suecia, Noruega y Suiza son los países que representan una menor mortalidad. Por el contrario, Irlanda y Dinamarca son los países con un mayor exceso de mortalidad. En cuanto a los hombres, nos encontramos con un patrón similar, con la excepción de Holanda que muestra una disminución de la mortalidad en comparación con sus países vecinos. Los países que muestran un exceso de mortalidad son más altas en Portugal e Irlanda.

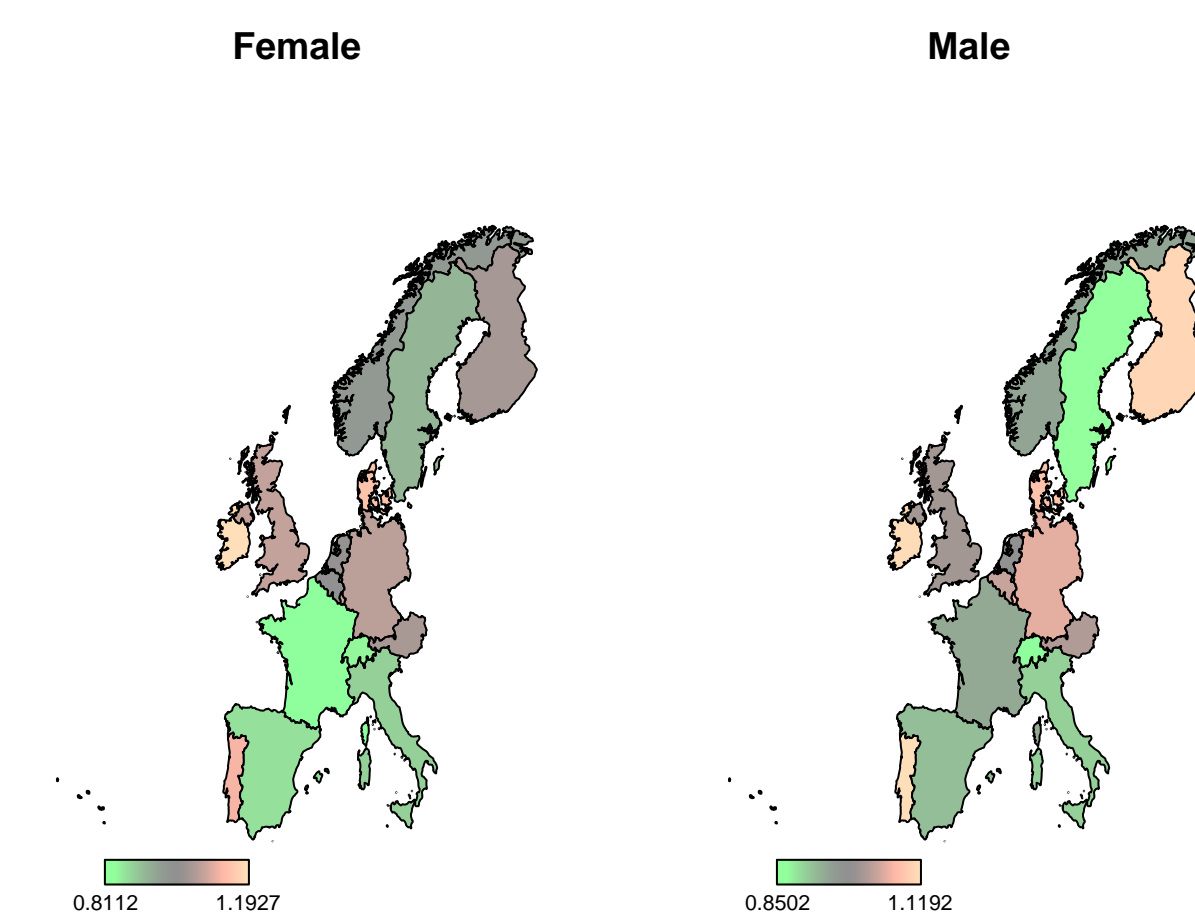


Figure 5: Efectos espaciales para ambos sexos

Como conclusión principal, podemos decir que hoy en día, las diferencias históricas en la mortalidad entre los hombres y las mujeres parecen estar desapareciendo, tendiendo a ser igual o incluso invertido en el sentido de que la mortalidad de las mujeres es más alta que la de los hombres. Esto es especialmente debido al aumento de la exposición de la primera a la intoxicación por la nicotina, el aumento de la aparición de tumores y otras enfermedades directamente relacionadas con este factor.

References

- *Debón, A., Molina, I., Cabrera, S. and Pellicer, A. (2013) Mathematical methodology to obtain and compare different embryo scores. *Mathematical and Computer Modelling*, 57(5), 1380-1394.
- *Debón, A. and Carlos Garcia-Díaz, J. (2012) Fault diagnosis and comparing risk for the steel coil manufacturing process using statistical models for binary data. *Reliability Engineering and System Safety*, 100, 102-114.
- *Ward, M. D., and Gleditsch, K. S. (2008) Spatial regression models *Sage*, (Vol. 155)
- *Team, R. C. (2005) R: A language and environment for statistical computing. *R foundation for Statistical Computing*.
- *Anselin, L. (2003) *GeoDaTM 0.9 user's guide* Urbana, 51, 61801.

Directora

Ana María Debón Aucejo. Titular de Universidad. Centro de Gestión de la Calidad y del Cambio. Universitat Politècnica de València. E-46022. Valencia. Spain. Tel: +34 963877007 (Ext. 74961); Fax: +34963877499; E-mail: andeau@cio.upv.es

Colaboradores

- Inma Molina Botella. Unitat de Reproducció, Hospital Universitari i Politènic "La Fe". Valencia. Spain.
- Francisco Montes Suay. Catedrático en el Departament d'Estadística i Investigació Operativa. Universitat de València. Spain.
- Adina Iftimi. Becaria FPI en el Departament d'Estadística i Investigació Operativa. Universitat de València. Spain.