



## Facultade de Administración e Dirección de Empresas

**Traballo de  
Fin de Grao**

Avaliación humana de  
modelos de aprendizaxe  
automática para  
mellorar o seu impacto  
na medicina

Sandra Pena Iglesia

07/2025

## Resumo

O presente Traballo de Fin de Grao propón a creación dun modelo de aprendizaxe automática para predicir o risco de nacementos de bebés con baixo peso. O modelo desenvolvido en Python partiu dunha base de datos do proxecto IMPACT-BCN, con variables demográficas e clínicas das nais, así como medidas dos bebés nas diferentes ecografías. Seleccionámolas en base a unha serie de criterios para lograr a maior precisión posible, alcanzando no mellor modelo un 73,13% de precisión .

De forma paralela ao desenvolvemento do modelo, planificamos a elaboración dunha enquisa para coñecer o grao de dificultade de comprensión da información presentada nos informes médicos das ecografías á vez que preguntamos polo nivel de aceptación da IA por parte das participantes. Observamos que son termos demasiado técnicos para dar unicamente con valores e que, o uso de rangos numéricos ou analoxías pódennles axudar a entender mellor os resultados. Por outra parte, a maioría mostran unha actitude receptiva ao uso da intelixencia artificial, pero si é certo que seguen preferindo un diagnóstico por parte dos profesionais médicos.

O estudo conclúe cun resumo da análise que se levou a cabo ao longo do proxecto e con aplicacións futuras.

Número de palabras (sen ter en conta a páxina inicial, o resumo, o índice, a bibliografía e os anexos): 10.629

# Índice

Resumo .....	2
Índice.....	3
Índice de táboas, gráficos ou figuras .....	5
Introdución.....	7
Marco teórico.....	9
Planificación .....	14
Estudo experimental .....	15
1. Hipótese .....	15
2. Obxectivos específicos .....	16
3. Metodoloxía .....	17
<b>3.1. Modelos predictivos</b> .....	17
3.1.1. Preparación e análise dos datos .....	17
3.1.2. Matriz de correlación .....	21
3.1.3. Construción e avaliación dos modelos.....	22
<b>3.2. Avaliación humana mediante enquisas</b> .....	31
3.2.1. Preparación da enquisa .....	31
3.2.2 Análise das respostas.....	34
Conclusións, limitacións e posibles ampliacións .....	56
Bibliografía .....	59
Anexo I .....	61
Enquisa.....	61
Anexo II .....	62
Código 1: modelo base de aprendizaxe automática.....	62
Código 2: modelo balanceado de aprendizaxe automática.....	64
Anexo III .....	67
Código 1: Relación entre perímetro abdominal e peso fetal estimado (semana 32) .....	67
Código 2: Matriz de correlación de Spearman (tipo para correlacións) .....	67
Código 3: Distribución do peso fetal na semana 28 .....	69
Código 4: Distribución da idade das participantes .....	70
Código 5: Nivel educativo das participantes (tipo dos diagramas de torta) .....	71
Código 6: Participantes con estudos do ámbito sanitario (tipo dos diagramas de barras) .....	71
Código 7: Comprensión de translucencia nucal en relación coa formación (tipo para boxplots) .....	72

Código 8: estatística T de Student .....	73
Código 9: Chi-cuadrado .....	73

## Índice de táboas, gráficos ou figuras

Figura 1. Planificación da parte empírica .....	14
Figura 2. Relación entre perímetro abdominal e peso fetal estimado (semana 32) .....	20
Figura 3. Matriz de correlación de Spearman.....	21
Figura 4. Validación StratifiedKfold .....	24
Figura 5. Distribución do peso fetal na semana 28.....	26
Figura 6. Resultados do modelo 5 .....	29
Figura 7. Matriz de correlación de Spearman coa semana 32 .....	30
Figura 8. Resultados do modelo 6 .....	30
Figura 9. Táboa comparativa dos modelos.....	31
Figura 10. Distribución da idade das participantes.....	34
Figura 11. Nivel educativo das participantes.....	35
Figura 12. Participantes con estudos do ámbito sanitario .....	36
Figura 13. ¿As participantes tiveron algún embarazo previo?.....	37
Figura 14. Grao de coñecemento do termo "translucencia nucal" .....	38
Figura 15. Capacidade de interpretar "translucencia nucal" a través de rangos numéricos .....	39
Figura 16. Capacidade de interpretar "translucencia nucal" a través de comparacións .....	40
Figura 17. Grao de comprensión de "diámetro biparietal" .....	41
Figura 18. Grao de comprensión do diámetro biparietal a través de rangos .....	42
Figura 19. Grao de comprensión do diámetro biparietal a través dunha comparación .....	43
Figura 20. Porcentaxe de participantes que se centrarían unicamente na conclusión.....	44
Figura 21. Porcentaxe de participantes que escoitaron falar dos percentís.....	45
Figura 22. Porcentaxe que sabe o que significa un percentil alto ou baixo .....	46
Figura 23. Respostas a pregunta de que significa que o bebé se encontra no p10 .....	46
Figura 24. Posibilidade de buscar todos os parámetros dos informes .....	47
Figura 25. ¿quedarían máis tranquilas coñecendo tódolos parámetros?.....	48
Figura 26. Como axudan os elementos cotiás a interpretación dos parámetros .....	48
Figura 27. Preferencia de formatos de presentación de informes .....	49
Figura 28. Aceptación dun modelo de IA con validación humana.....	50
Figura 29. Preferencia de diagnósticos levados a cabo por médicos.....	50
Figura 30. Grao de confianza nun modelo de IA para predicir baixo peso .....	51
Figura 31. Necesidade de información adicional do modelo de predición de baixo peso .....	51
Figura 32. Grao de aceptación da IA na medicina .....	52
Figura 33. Comprensión de translucencia nucal en relación coa formación .....	53

Figura 34. Comprensión do diámetro biparietal en relación coa formación .....53  
Figura 35. Táboa de continxencia.....55

## Introdución

Actualmente, o uso de intelixencia artificial (IA) e de algoritmos na medicina axudan a analizar grandes conxuntos de datos. Os sistemas de IA son unha ferramenta de apoio que axuda a mellorar a precisión dos diagnósticos e a eficiencia no traballo aos profesionais médicos. Sen embargo, a inclusión de estes métodos na práctica clínica está sendo un procedemento lento xa que moitos sanitarios non teñen os coñecementos necesarios e tampouco existe unha validación real destas ferramentas na parte práctica (Vega, 2020).

No ámbito da obstetricia, un dos estudos levados a cabo consiste no uso da aprendizaxe automática para a predición de nacementos con baixo peso para a idade xestacional (Saw,2021). Neste contexto, o presente traballo propón a elaboración dun modelo de aprendizaxe automática para, baseándose en variables maternas e medidas fetais, saber se un bebé vai nacer con baixo peso, seguindo a liña de estudo de Saw (2021). Por outra parte, tamén é importante facer unha análise da aceptación da IA no uso práctico (Vega,2020), polo que este traballo se completou cun experimento de validación humana baseada en enquisas. Estas céntranse no rendemento do modelo, pero tamén no grao de comprensión dos informes presentados nas ecografías e na aceptación das explicacións xeradas pola IA na obstetricia.

En concreto, elaboramos un modelo con Python para predicir se un bebé pode nacer con baixo peso ou baixo peso severo. Empregamos unha base de datos con variables demográficas e clínicas da nai, así como das métricas das ecografías obtidas no primeiro, segundo e terceiro trimestre. As variables seleccionámolas segundo estudos previos e en base á relevancia clínica para a predición.

Na segunda parte, analizamos as avaliacións baseadas no ser humano, entendidas como aquelas que analizan persoas non expertas e que teñen a necesidade de comprender conceptos técnicos. Neste caso analizamos as opinións de mulleres embarazadas como usuarias finais sobre a comprensión dos parámetros que se amosan nos informes das ecografías durante o embarazo.O obxectivo é coñecer se entenden os conceptos que aparecen nos informes médicos derivados das ecografías. Incluímos os parámetros que poden provocar un aumento do risco dun nacemento con baixo peso. Centrámomos tamén en entender variables técnicas dos informes e saber cal sería o seu valor nunha situación de crecemento normal. Para iso, tivemos en conta documentos reais con datos de ecografías do primeiro, segundo e terceiro trimestre, así como as medidas tras o nacemento.

Mediante a realización de enquisas, as participantes valoraron o distinto nivel de comprensión da presentación de resultados en informes médicos mediante o uso de imaxes, analogías ou gráficas. Tamén avaliamos o nivel de aceptación da intelixencia artificial por parte das pacientes e se estarían dispostas a obter explicacións xeradas por esta tecnoloxía.

## Marco teórico

### 1. Intelixencia Artificial e a aplicación en medicina

A intelixencia artificial (IA) son as tecnoloxías que permite aos dispositivos entender e traducir a linguaxe do ser humano, tanto falada como escrita, analizar información ou facer recomendacións, entre outras. Este campo é moi amplo porque se nutre de ramas como as matemáticas, a programación, lingüística ou incluso a filosofía e psicoloxía (aínda que son áreas con máis dificultade de aplicación). Para que se poida levar a cabo a aprendizaxe é necesario o uso de algoritmos. Estes son un conxunto de pasos para realizar unha acción concreta e analizar os resultados, así como determinar as posibles melloras<sup>1</sup>. Esta tecnoloxía está composta por unha ampla variedade de campos entre os que podemos encontrar a aprendizaxe automática e a profunda. A primeira analiza as relacións que hai entre os datos dun conxunto e identifica patróns para facer predicións ou clasificacións, baseándose nesa información. A segunda utiliza redes neuronais para extraer parámetros e características directamente de bases de datos moi grandes.<sup>2</sup>

A tecnoloxía e a medicina estanse desenvolvendo de forma conxunta nas últimas décadas. Estes avances tecnolóxicos están modificando a rama da saúde e as necesidades que van da man, con máis esixencia das novas tecnoloxías. A IA neste ámbito xoga un papel importante porque pode axudar a tomar decisións analizando información e adestrando algoritmos a partir dela. O obxectivo é que se poida aplicar en casos concretos, aprendendo da realidade e das normas xerais (Avila-Tomás,2020).

O uso da IA na medicina ten como meta poder procesar información médica, apoiando aos profesionais e axudándolles a mellorar nos seus diagnósticos. A aprendizaxe automática permite analizar datos de historias clínicas electrónicas (EHRs) como imaxes ou textos que almacenan

---

<sup>1</sup> ¿Qué es la inteligencia artificial o IA? | Google Cloud. (s. f.). Google Cloud. <https://cloud.google.com/learn/what-is-artificial-intelligence?hl=es-419>

<sup>2</sup> Martínez, E. (2025, 17 marzo). Differences between Artificial Intelligence, Machine Learning and Deep Learning - ATRIA Innovation. *ATRIA Innovation*. <https://atriainnovation.com/en/blog/differences-between-artificial-intelligence-machine-learning-and-deep-learning/>

información médica dos pacientes (Gambetti, 2025). Pode servir como ferramenta para facilitar a detección de enfermidades, así como para mellorar a experiencia dos pacientes<sup>3</sup>. Ademais, a evolución da sociedade obriga aos médicos a empregar esta ferramenta debido aos beneficios que presenta tanto para os médicos como para os pacientes. Os sanitarios verán a súa carga administrativa reducida, ademais de contar cun apoio na toma de decisións. Por outra banda, o impacto para a poboación será positiva polo incremento da seguridade nos diagnósticos médicos. Ademais, podería chegar a ser unha oportunidade para mellorar a atención médica en zonas de difícil acceso (Lanzagorta-Ortega, 2022).

Aínda que sabemos que ten moito potencial, esta tecnoloxía ten un uso real menor ao esperado. Unha das causas principais é a falta de confianza por parte dos sanitarios nas respostas da IA, principalmente se descoñecen o que hai detrás do modelo, é dicir, sobre que información traballa para obter as predicións. Ante este problema xorde a Intelixencia Artificial Explicable (XAI) que ten como obxectivo mostrar respostas transparentes aumentando así a confianza no modelo. Así, os profesionais entenderían por que a IA mostra unha resposta concreta e favorecería a adopción no ámbito médico. (Gambetti, 2025).

Hai tres clasificacións para avaliar a XAI:

- Baseadas na aplicación: implica probas con situacións do mundo real para comprobar a facilidade de uso da IA.
- Baseadas no ser humano: tarefas simples con usuarios non expertos para saber se comprenden as explicacións con claridade.
- Baseadas na avaliación por proxy: empréganse métricas automatizadas para estimar a capacidade desta tecnoloxía, medindo a calidade das súas explicacións sen ter en conta ao ser humano (Gambetti, 2025).

## 2. Baixo peso ao nacer

---

<sup>3</sup> ¿Qué es la inteligencia artificial en medicina? (2024, octubre 14) <https://www.ibm.com/es-es/topics/artificial-intelligence-medicine>

A día de hoxe, segundo a Organización Mundial da Saúde, a definición de baixo peso ao nacer é cando está por debaixo de 2500g e estímase que o sofren entre un 15% e un 20% dos bebés do mundo<sup>4</sup>. Esta é unha das principais causas de mortalidade prenatal<sup>5</sup>. Para referirse a este termo tamén se utiliza o concepto de PEG que significa “pequeno para a súa idade xestacional” (Grisolía,2008).

Para saber se un feto é máis pequeno do normal para a idade que ten utilízase un ultrason. Débese normalmente a unha restrición do crecemento intrauterino e que se define como o peso que está por debaixo do percentil 10. As causas poden incluír enfermidades xenéticas, metabólicas hereditarias, anomalías cromosómicas, embarazo múltiple, placenta anormal ou a presión arterial alta (Osuchukwu,2020). Hai que ter en conta que cando os bebés nacen con baixo peso, en moitas ocasións, presentan un crecemento recuperador ou “catch-out” que é o aumento da velocidade de crecemento para a idade e sexo do neonato durante un tempo, tras un escaso crecemento (Grisolía, 2008).

Algún estudo analiza a relación que hai entre a diabetes xestacional con este tipo de nacementos. Cando se detecta, a solución é unha dieta para a nai que modifica o incremento de peso as mulleres. Se o seu peso non aumenta da maneira esperada e é máis limitado pode provocar nacementos con baixo peso para a idade xestacional. En numerosos estudos atoparon esta relación, aínda que non está claro se está é a causa razón principal (Modino,2019).

Xa fai anos estaba determinado que había unha series de causas que eran as que podían levar a este peso inferior. Podemos encontrar: causas fetais: cromosomopatías, anomalías conxénitas ou enfermidades xenéticas. Tamén causas maternas: enfermidades médicas como hipertensión arterial, enfermidade renal, diabetes, infeccións, estado nutricional ou abuso de substancias.. Tamén se incluían outras como causas útero-placentarias ou causas demográficas (idade

---

<sup>4</sup> Metas mundiales de nutrición 2025: documento normativo sobre baixo peso al nacer. 30 de diciembre de 2014). Who.int; World Health Organization. <https://www.who.int/es/publications/i/item/WHO-NMH-NHD-14.5>

<sup>5</sup> Safety, N. A. F. (s. f.). *Metas mundiales de nutrición 2025: documento normativo sobre baixo peso al nacer.* <https://www.who.int/es/publications/i/item/WHO-NMH-NHD-14.5>

materna extrema, etnia, nuliparidade(muller que nunca tivo un parto<sup>6</sup>), fillo previo nacido con baixo peso...)(Grisolía, 2008).

### 3. Aplicacións da IA no SGA.

O uso actual da IA neste ámbito é bastante reducido ao igual que a aplicación da aprendizaxe automática. Isto débese a que hai escaseza de bases de datos grandes para poder analizar e predicir variables neste campo. A limitación de mostras fai que a tecnoloxía tampouco se poida aplicar correctamente, dificultando a xeneralización e por tanto, provocando inxustizas nas diferentes sociedades, que se producen cando, por exemplo, un modelo funciona sempre peor para unha determinada subpoboación pouco representada no conxunto de adestramento.

Sen embargo, hai algún estudo de aplicación da tecnoloxía para determinar a idade xestacional do bebé mediante medicións barométricas fetais (cabeza, abdome e fémur). A partir de aquí Lee et al, 2022, propuxo tres modelos principais: un de imaxes nun plano estándar, un de vídeo e un último que funcionaba como unha combinación entre os dous primeiros. A nivel estatístico, calquera dos tres foi superior as estimacións do peso coas ecografías tradicionais (Lee, 2022).

No artigo de Saw et al., 2021, os investigadores utilizaron tres modelos de aprendizaxe automática para determinar se o bebé ten posibilidades de nacer con baixo peso ou baixo peso severo con parámetros do segundo trimestre. Os algoritmos utilizados foron RandomForest, Máquina de Vectores de Soporte e Perceptrón Multicapa. Algunha das variables que tiveron en conta foi a idade materna, o peso fetal estimado, a circunferencia abdominal ou o índice de pulsatilidade (IP) uterino que é un parámetro utilizado nas ecografías para avaliar a circulación fetal(Rivas, 2016). Tamén a lonxitude do fémur, grosor do prego nucal ou o diámetro do cerebro transversal. Dividiron os datos aleatoriamente en conxunto de adestramento e validación ( 80%) e para a proba (20%). Lograron alcanzar unha precisión do 70% ao 73% na predición, e o IP uterino e o grosor do prego nucal foron preditores moi importantes (Saw, 2021).

---

<sup>6</sup> *Qué es el término nulípara. Diccionario médico. Clínica U. Navarra. (s. f.).* <https://www.cun.es>. <https://www.cun.es/diccionario-medico/terminos/nulipara>

A día de hoxe considerase que a IA pode mellorar o coidado prenatal e identificar cando a nai ten risco de que o bebé teña un peso inferior ao normal, así como identificar complicacións durante o embarazo ou o parto. Outra das funcións é saber se pode ter anomalías ou trastornos grazas a probas como as ecografías (Wen, 2022).

Por último, como tratamos o tema da intelixencia artificial na medicina tamén é importante ter en conta que vai asociada a desafíos e riscos. Unha das preguntas que máis escoitamos dende que a IA tivo un pico no seu desenvolvemento é se chegará na substituír o traballo do ser humano. Aquí é onde entra a dúbida de se nalgún momento esta tecnoloxía chegará traballar polos médicos ou por parte deles. A resposta é clara: non. En primeiro lugar, é imposible automatizar toda as tarefas que trae consigo a atención clínica. Un exemplo son as decisións baseadas en diagnósticos diferencias que teñen que ter en conta a necesidade de mostrar empatía e sentimento e de razoar en situacións complicadas, algo que unicamente os seres humanos son capaces de facer. Tamén debemos recordar que moitas veces os modelos de IA non son explicables e non sabemos que pasa por detrás e como obteñen os datos. Por todo isto, a responsabilidade dos médicos é tan importante e son eles os que deben tomar as decisións finais (García-López, 2023).

Por outra parte, existen as consideracións éticas que se deben de ter en conta ao aplicar esta ferramentas no ámbito da saúde. Para empezar, a evidencia que analizaron é que os sistemas de IA poden chegar a presentar prácticas discriminatorias. En moitos países chegaron a facer manifestos e guías para determinar como se debe implantar, centrándose en propoñer códigos de conducta que funcionen como guías de uso, tendo en conta a parte moral. A nivel de modelos para o ámbito sanitario o problema céntrase máis nos deseñadores e usuarios que no propio modelo. Hai que ter en conta que se trata dun enorme volume de datos e cunha gran sensibilidade. Por iso, cando se utilice os pacientes deberían saber se se vai aplicar aprendizaxe automático sobre o seu historial e os sanitarios teñen que ser capaces de entender como se desenvolve este modelo e a súa precisión. Cando se entrena cunha poboación que non representa ao enfermo, pode xerar unha recomendación inválida. Os médicos teñen que ser capaces de determinar cal é a decisión final e por que lle fan caso á IA ou por que non. Por tanto, os principios básicos que deben cumprir a nivel ético son: autonomía, beneficencia, non maleficencia e xustiza. En resumo, a ética neste campo, polo momento, consiste en utilizar aprendizaxe automática como unha ferramenta de apoio para a toma de decisións, pero mantendo a responsabilidade dos médicos (García-López, 2023).

## Planificación

A planificación estruturouse en varias fases que se observan no diagrama de Gantt da Figura 1. Este céntrase na enquisa porque é a que esixiu máis tempo para a súa elaboración.

Tal e como se observa abarcou un amplo período de tempo e dividiuse en subfases: formulación de hipóteses, definición de perfil das participantes, métodos de difusión, tamaño da mostra e redacción da mesma (con borradores iniciais ata obter a versión final). A fase de difusión e recollida de respostas foi a que ocupou máis tempo xa que se trata dun tema moi específico, limitando a colaboración na mesma.

Figura 1. Planificación da parte empírica



Elaboración propia

Por outra banda, o desenvolvemento do modelo en Python non figura explicitamente na planificación temporal porque o fixemos de forma paralela á enquisa, pero de forma máis flexible xa que partimos dunha base de datos feita e procesada.

Por tanto, a planificación permitiu garantir o desenvolvemento da parte máis complexa e sensible: a avaliación humana a través de enquisas. Esta parte esixiu máis tempo e máis carga de traballo pola elaboración, desenvolvemento e análise das respostas.

## Estudo experimental

### 1. Hipótese

Segundo estudos previos, o formato de presentación da información inflúe na comprensión dos datos por parte dos pacientes. Un estudo realizado por investigadores de Weill Cornell Medicine, fai 5 anos, aseguraba que o uso de analogías melloraba a interpretación dos termos máis técnicos, incluso por encima de imaxes ou gráficas<sup>7</sup>. Así mesmo, no estudo de Armero, 2023, elaboraron enquisas anónimas dirixidas a xestantes sobre o uso da IA e os seus beneficios. As conclusións finais mostran que esta tecnoloxía é positiva para axudarlles aos sanitarios, pero que é necesaria a parte humana, cun médico capaz de crear unha relación próxima co paciente. Como conclusión, asegura que se trata dun soporte para axudarlles, pero non é unha ferramenta que poida substituílos (Armero, 2023).

En base a estes dous estudos, podemos desenvolver as seguintes hipóteses:

1. O uso de enquisas permite coñecer o nivel de aceptación que teñen as mulleres embarazadas das metodoloxías baseadas na intelixencia artificial para a predición de nacementos con baixo peso. Desta forma podemos analizar se confían na IA, se a consideran unha ferramenta de apoio e tamén a importancia que lle dan ao papel dos médicos.
2. As enquisas permiten saber cal é o grao de coñecemento das embarazadas con respecto aos parámetros que aparecen nos informes das ecografías. Tamén determinan cal é o formato preferido polas xestantes para presentar dita información: a través de imaxes, comparacións, frases curtas ou rangos numéricos. O obxectivo é avaliar a

---

<sup>7</sup> (Information visualization helps patients contextualize patient. (s. f.). Weill Cornell Medicine. <https://phs.weill.cornell.edu/news/information-visualization-helps-patients-contextualize-patient-reported-outcomes#:~:text=Visual%20analogies%20have%20shown%20to,by%20Weill%20Cornell%20Medicine%20investigators> )

claridade da presentación da información e identificar cales son os métodos que melloran a comprensión destes conceptos técnicos.

Tamén buscamos saber, a través dunha perspectiva empírica centrada nas xestantes, se aceptan o uso da aprendizaxe automática para os diagnósticos ou se prefiren explicacións unicamente dos profesionais médicos.

## **2.Obxectivos específicos**

O principal obxectivo é estudar como o uso da IA pode axudar as nais embarazadas a comprender parámetros fetais, así como a súa aplicación en caso de nacementos de bebés de baixo peso.

Os obxectivos son:

1. Avaliar a utilidade da IA para entender as analíticas fetais por parte das nais : un dos principais obxectivos é comprender ata que punto as xestantes confían na información proporcionada por sistemas de IA para os diagnósticos e análises prenatais como a falta de crecemento fetal ou a probabilidade de nacer con baixo peso. É importante saber se as nais son capaces de interpretar correctamente parámetros como os percentís fetais ou a hemoglobina.
2. Analizar a aceptación dos modelos predictivos, así como a de confianza que teñen neles: Identificamos se esta nova tecnoloxía xera confianza e se prefiren que os diagnósticos sexan unicamente dos profesionais. É importante coñecer a relación entre a utilidade da intelixencia artificial, da claridade das explicacións e das preferencias das xestantes. Tamén tivemos en conta as variables socio demográficas, é dicir, os segmentos en base a idade, á educación ou o país de nacemento.
3. Comparar a efectividade dos distintos métodos de explicación (técnico, visual, situacións reais...): impacto da presentación das explicacións dos resultados, é dicir, se as técnicas tradicionais funcionan mellor que o uso de visualizacións ou comparacións con situacións da vida real. Buscamos coñecer se hai algunha diferenza na comprensión ao utilizar formatos visuais, incluso, gráficos.

## 3. Metodoloxía

### 3.1. Modelos predictivos

#### 3.1.1. Preparación e análise dos datos

Nesta parte xeramos o modelo de aprendizaxe automática para determinar, en base a características do feto e da nai, se o bebé nacerá con baixo peso (SCGA) ou baixo peso severo (SGA severo). Para iso utilizouse unha base de datos que nace a partir do ensaio clínico IMPACT-BCN (Improving for a Better Prenatal Care Trial Barcelona), desenvolvida no BCNatal (Hospital Clínic e Hospital Sant Joan de Déu, Barcelona) dende o 2017 ata o 2019. Este estudo forma parte do proxecto COLLAGE (Comprehensive Machine Learning- based Analysis of Fetal Growth Trajectories), co-dirixido pola Universidade de Santiago de Compostela e a Universitat Pompeu Fabra de Barcelona. O acceso a esta información parte dun acordo de transferencia de datos entre o hospital e a USC, garantindo a confidencialidade da información clínica. Esta base de datos consta de 520 variables e 746 embarazadas. Inclúense datos demográficos das mulleres, antecedentes de enfermidades, controis que se levan a cabo durante o embarazo, medidas do crecemento nas ecografías en distintos trimestres e os resultados finais.

Para poder traballar cos datos importouse a librería de pandas e numpy. Estas permiten a manipulación dos datos. Lemos o arquivo de Excel coa base de datos nomeada anteriormente.

As variables incluídas no modelo inicial son:

Datos xerais da nai:

- **Maternal\_age:** idade da nai no momento do embarazo.
- **Nuliparity:** se a nai tivo algún embarazo anteriormente. As respostas son de si ou non e convertémolas a 1 e 0 respectivamente.
- **Ethnicity:** Etnia da nai. As opcións son negra, asiática, india, latina, magrebí ou branca. Estas codificámolas do 0 ao 5.
- **Class\_socioec:** clase socio económica. As respostas podían ser: baixa, media e alta. Transformáronse en numéricas do 0 ata o 2, de menor a maior.

Enfermidades:

- **HTAcronic:** Se a nai foi diagnosticada de hipertensión arterial crónica.

- **PE:** É a preeclampsia, é dicir, o aumento da presión arterial a partir da semana 20 da xestación<sup>8</sup>.
- **GDM:** diagnosticada de diabetes xestacional. As opcións eran si e non e codificámolos con 1 e 0 respectivamente.

Variabes para a semana 20:

- **DBP20s:** Diámetro biparietal. Esta medida é utilizada na ecografía obstetricia para avaliar o crecemento e o desenvolvemento do feto durante o embarazo. Correspóndese ca distancia que hai entre os dous osos parietais do cranio fetal, medida no punto máis amplo. Na semana 20 debería situarse entre 43 mm e 50 mm<sup>9</sup>.
- **PC20s:** perímetro cefálico. É a distancia da cabeza dun neno na súa parte máis ampla, dende a parte derriba das células ata a parte posterior da cabeza<sup>10</sup>.
- **PA20s:** perímetro abdominal. É o contorno do abdome, lixeiramente por encima do embigo<sup>11</sup>.
- **LF20s:** lonxitude do fémur.

Variabes da semana 28:

- **DBP28s:** diámetro biparietal na semana 28.
- **PC28s:** perímetro cefálico na semana 28.
- **PA28s:** perímetro abdominal na semana 28.
- **LF28s:** lonxitude del fémur na semana 28.

Variabes a predicir :

---

<sup>8</sup> Preeclampsia en el embarazo: síntomas y cuidados | CUN. (s. f.). <https://www.cun.es>. <https://www.cun.es/chequeos-salud/embarazo/preeclampsia#:~:text=La%20preeclampsia%20es%20una%20enfermedad,de%20prote%C3%ADnas%20por%20la%20orina>.

<sup>9</sup> Diámetro biparietal: definición médica. Diccionario CUN. (s. f.). <https://www.cun.es>. <https://www.cun.es/diccionario-medico/terminos/diametro-biparietal>

<sup>10</sup> Perímetro cefálico. (s/f). Medlineplus.gov. Recuperado el 4 de julio de 2025, de <https://medlineplus.gov/spanish/ency/article/002379.htm>

<sup>11</sup> Natalben. (2024, 13 noviembre). *Ecografía: Medición del perímetro abdominal*. Natalben. <https://www.natalben.com/ecografia-embarazo/tercer-trimestre-medicion-perimetro-abdomen-fetal>

- **SGA\_birth (<p10):** un dos resultados obxectivo que indica se o bebé nace con peso menor ao percentil 10 (isto indica que o bebé pesa menos que 9 de cada 10 bebés da mesma idade xestacional<sup>12</sup>). O número 1 corresponde ca resposta afirmativa e o 0 coa negativa.
- **SeverageSGA (<p3):** a outra predición que fai o clasificador. Indica se o bebé nace con baixo peso severo, que se corresponde co percentil 3 (valor por debaixo do cal se atopa o 3% da poboación con menor peso<sup>13</sup>). Nesta situación estamos ante un caso de peso extremadamente baixo que require unha vixilancia especial durante o embarazo<sup>14</sup>.

No Excel preséntase outra variable importante: o peso fetal estimado. A pesar de estar directamente relacionado co peso baixo ao nacer non o tivemos en conta no modelo porque podería traer problemas consigo. Por unha parte, xa é parte do resultado obxectivo e reduce a capacidade do modelo de aprender das variables que realmente nos interesan e por outro lado, está demasiado correlacionado co resultado e podería ignorar os outros factores. O obxectivo é asegurar que sexa un modelo máis global e non dependa, en gran medida, dunha única variable. De feito, na Figura 2 pódese ver a forte correlación positiva que existe entre o peso estimado e o perímetro abdominal. Observamos que a medida que unha aumenta, a outra tamén o fai. O mesmo ocorre cas outras métricas de crecemento do bebé. Descartámola a causa disto, para evitar que estean altamente relacionadas entre si e evitando unha posible multicolineariedade.

---

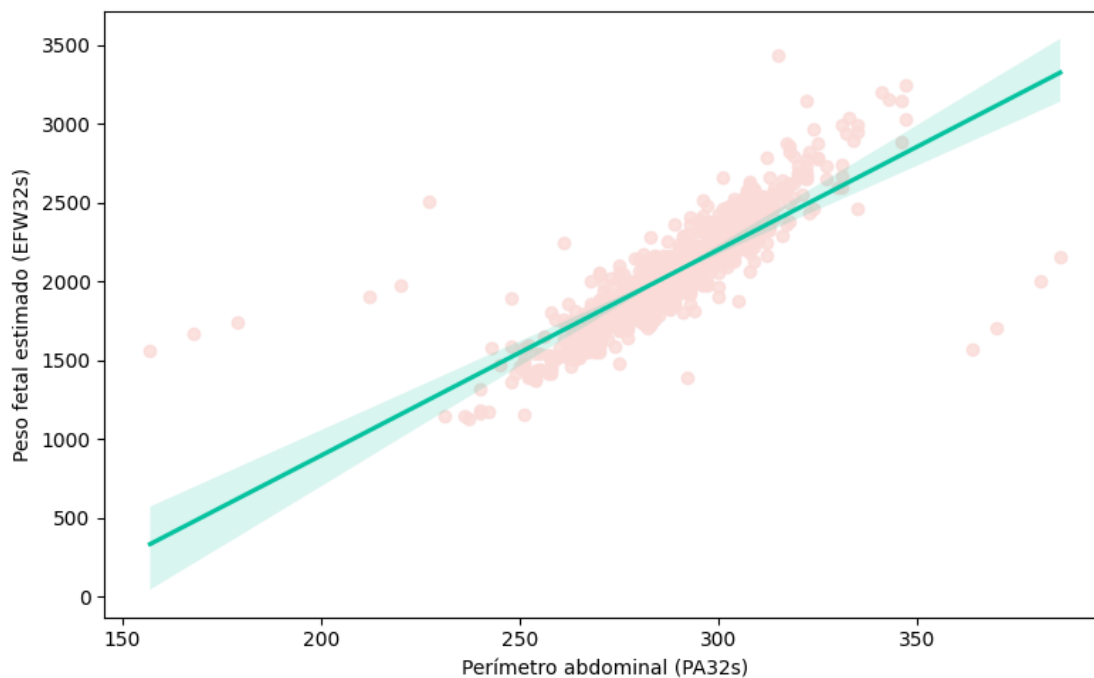
<sup>12</sup> *Fetal Growth Restriction (FGR)*. (s. f.). Stanford Medicine Children's Health.  
<https://www.stanfordchildrens.org/es/topic/default?id=fetal-growth-restriction-fgr-90-P05572#:~:text=Es%20frecuente%20que%20se%20describa,peque%C3%B1os%20para%20a%20edad%20gestacional%22>.

<sup>13</sup> *Pequeño para edad gestacional: Talla baja de adulto y tratamiento con hormona de crecimiento*. (s. f.). Quirónsalud.  
<https://www.quironsalud.com/es/comunicacion/contenidos-salud/pequeno-edad-gestacional-tratamiento-talla-baja-hormona-cre#:~:text=En%20el%20contexto%20del%20crecimiento,retardo%20en%20el%20crecimiento%20intrauterino>.

<sup>14</sup> García, D. E. (2024, 13 noviembre). *Cómo interpretar los percentiles del bebé por ecografía*. Natalben.  
<https://www.natalben.com/el-embarazo-y-tus-dudas/embarazo-bebe-percentiles-interpretar-ecografia#:~:text=Cuando%20es%20igual%20o%20menor,los%20percentiles%20varios%20a%C3%B1os%20m%C3%A1s>.

A maioría dos puntos están agrupados na liña de regresión e, aínda que hai puntos dispersos, estes poderíanse explicar por outras variacións en factores da base de datos como as condicións maternas.

Figura 2. Relación entre perímetro abdominal e peso fetal estimado (semana 32)



Elaboración propia

Na elaboración deste modelo predictivo tamén é fundamental explicar por que excluimos outras variables dispoñibles na base de datos. Os criterios empregados para eliminar as variables foron:

- Pouca correlación entre a variable e as dúas obxectivo: SGA\_bith (<p10) e severeSGA (<p3). Consideramos que podían aportar ruído e reducir a precisión do modelo.
- Problemas ca calidade dos datos. Algunhas contiñan un número elevado de valores faltantes que esixía ter que eliminar a maior parte das observacións.
- Outras non amosaban unha gran relevancia no modelo nin había evidencia en artigos científicos de que tiveran impacto no crecemento fetal.
- O seguinte paso foi fundamental para o funcionamento do modelo: transformar as variables categóricas en numéricas para que os algoritmos as procesaran directamente. Neste caso realizámola de forma manual, tendo en conta as categorías de respostas presentes no conxunto de datos. Podemos dividilas en tres grupos principais:
  1. As variables binarias. Estas son as que teñen a opción de “yes” ou “no”. Pasámolas directamente a valores de 1 e 0, respectivamente. Desta forma, o clasificador entende

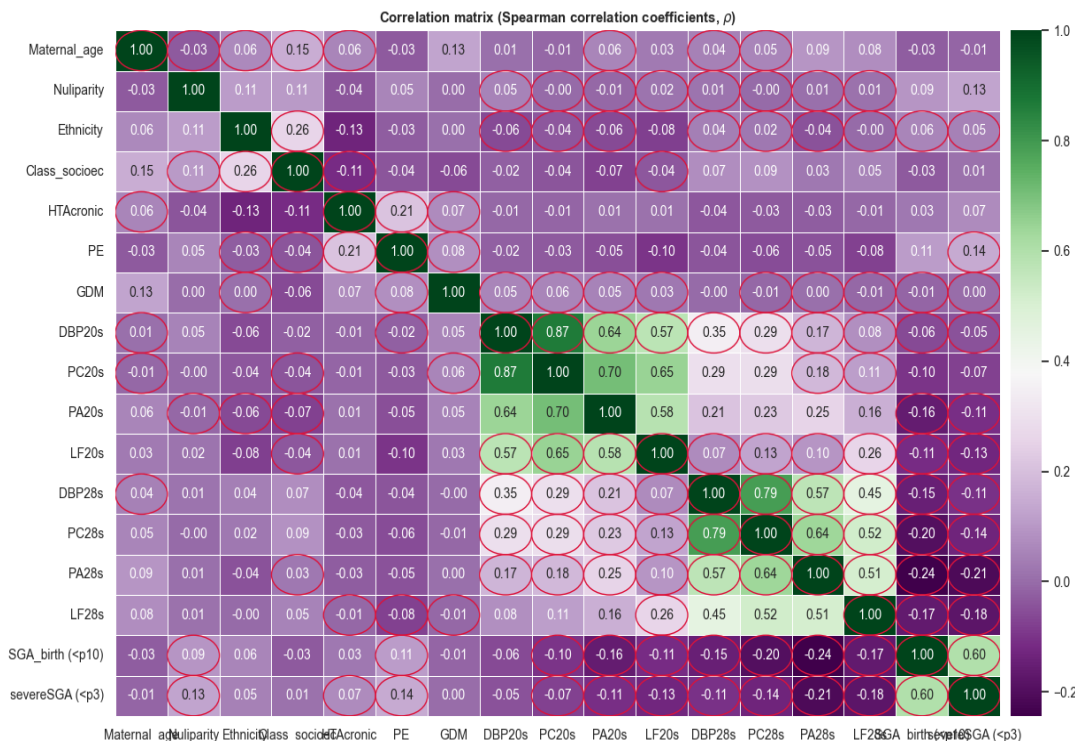
se hai presenza ou ausencia a través de números. Aplicouse as variables de “Nuliparity”, “HTAcronic”, “GDM”, “PE”, “SGA\_birth (<p10) e severeSGA(<p3).

2. A variable socioeconómica. Esta distribúese en tres categorías principais: “low”, “medium” e “high”. Codificouse do 0 ao 2, respectivamente.
3. A variable nominal de etnicidade. Esta non ten unha orde concreta, por iso se define como nominal. Os seus valores son “afroam”, “asian”, “indi”, “latin”, “magrebi” e “white”. Transformáronse, seguindo esa orde, do 0 ao 5. Cabe destacar que non existe ningunha orde, a diferenza do caso anterior que era por niveis e de forma ascendente.

### 3.1.2. Matriz de correlación

A través de Python tamén elaboramos unha matriz de correlacións, presentada na figura 3, para saber que variables tiñan maior relevancia. O número que se amosa nas celas da táboa son o coeficiente de correlación (rho).

Figura 3. Matriz de correlación de Spearman



Elaboración propia

A matriz da Figura 3 mostra a correlación cruzada entre todas as variables clínicas e demográficas das nais e cos resultados de nacemento con baixo peso ou con baixo peso severo, SGA\_bith (<p10) e severeSGA(<3), respectivamente.

O coeficiente de correlación de Spearman (ρ) é unha medida que indica o grado de relación entre dous conxuntos de datos. Móvense entre valores de -1 e 1, onde 0 implica ausencia de

correlación. A correlación de  $-1$  implica unha correlación negativa exacta e significa que ao aumentar "x", diminúe na mesma proporción "y". Cando a correlación é igual a 1 é unha correlación positiva exacta, o que quere dicir que aumenta "x" e aumenta de igual forma "y".

Na Figura 3, os números correspóndense co coeficiente de correlación,  $\rho$ , e os círculos vermellos están relacionados co nivel de significación estatística, é dicir, co valor p do test. A hipótese nula é que as variables non teñen correlación. O valor p é o que decide se se rexeita esta hipótese.

<sup>15</sup>O que se fai nestas probas é fixar inicialmente o nivel de significación (normalmente de 0,05). Se ao calculalo este é inferior rexéitase a hipótese nula <sup>16</sup>. Na Figura, un círculo vermello significa  $p < 0,05$  e que polo tanto, o nivel de correlación pódese considerar significativo estatisticamente.

Nesta matriz centrámonos principalmente nas dúas últimas columnas porque representan o peso por debaixo do percentil 10 e do percentil 3, é dicir SGA e SGA severo. As que mostran maior correlación (neste caso negativa) son as seguintes. PA28s, un perímetro menor durante a semana 28 aumenta o risco de que o bebé naza con baixo peso. É a que presenta valores máis próximos ao  $-1$ , sendo por tanto a que ten maior correlación. Ocorre o mesmo ca lonxitude femoral, o perímetro cefálico e o diámetro biparietal nese período. Canto máis pequenos, máis probable é que o bebé teña un peso considerado baixo.

Nuliparity ten unha correlación positiva, pero tamén baixa (de 0,13). Algunhas con valores inferiores son HTAcronic, ethnicity, maternal\_age, e class\_socioec.

Tamén teñen máis influencia as medidas do bebé na semana 28 que na semana 20. Podería indicar que para a predición teñen máis poder as das semanas máis próximas ao nacemento.

Por último, destacar que a diabetes xestacional presenta ausencia de correlación co peso.

Grazas á matriz tamén identificamos variables que teñen máis relación cos nacementos con pesos baixos e tivémolas en conta para facer os modelos.

### 3.1.3. Construción e avaliación dos modelos

---

<sup>15</sup> T-Test, Chi-Square, ANOVA, Regression, Correlation. . . (s. f.). <https://datatab.es/tutorial/correlation>

<sup>16</sup> T-Test, Chi-Square, ANOVA, Regression, Correlation. . . (s. f.). <https://datatab.es/tutorial/correlation>

## MODELO 1. MODELO BÁSICO

Para crear o modelo inicial dividimos as variables en dous conxuntos: independentes (almacenadas na variable  $X$ ) e as dependentes (se o bebé ten baixo peso ou non) almacenadas na variable  $y$ .

Ademais, normalizamos todas as características ao rango  $[0,1]$  porque así evitamos que as variables de maior escala dominen sobre as outras. Deste xeito, todas se moven nos mesmos niveis, mellorando a precisión do modelo. Neste caso utilizouse o normalizador MinMax que escala e traduce cada característica individualmente de modo que estea dentro do rango que desexamos (de 0 a 1) no conxunto de adestramento. Non reduce o efecto dos valores atípicos, se non que diminúe de forma lineal aos valores definidos, onde o punto de datos máis grande corresponde co valor máximo e o máis pequeno co mínimo<sup>17</sup>. Neste caso é importante levalo a cabo porque se traballa con datos con escalas moi diferentes como a idade materna ou a presión arterial.

O modelo seleccionado para a tarefa foi regresión loxística. No código ten a función de predicir se o bebé nacerá con peso inferior ao percentil 10 (baixo peso, SGA), a partir das variables seleccionadas anteriormente. A regresión loxística é un método estatístico que predí clases binarias. A variable obxectivo só ten dúas posibles clases, é dicir, é dicotómica. Esta regresión é un caso especial da lineal na que a variable a predicir é de natureza categórica, e que obtén a saída cunha función logit<sup>18</sup>.

Para a avaliación do modelo utilizouse un esquema de validación cruzada k-fold que é unha técnica habitual para estimar a capacidade de xeneralización dun modelo predictivo e mitigar o risco do sobreaxuste que se produce cando un modelo funciona moi ben nos datos de adestramento pero mal nos datos de test. Consiste en dividir o conxunto de datos en  $k$  particións (*folds*) e, en cada iteración, adestrar o modelo con  $k-1$  particións y avalialo coa restante. O

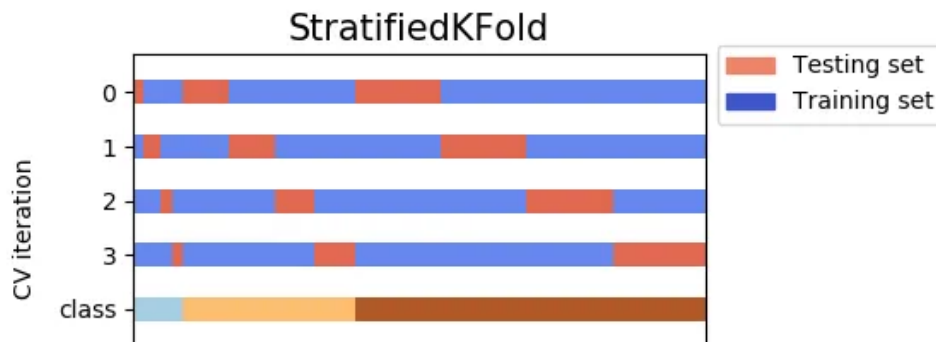
---

<sup>17</sup> *MinMaxScaler*. (s. f.). Scikit-learn. <https://scikit-learn.org/stable/modules/generated/sklearn.preprocessing.MinMaxScaler.html>

<sup>18</sup> Avinash Navlani. (2024, April 25). *Comprender la regresión logística en el tutorial de Python*. Datacamp.com; DataCamp. <https://www.datacamp.com/es/tutorial/understanding-logistic-regression-python>

procedemento repítese ata que todas as particións sexan utilizadas como conxunto de validación. A métrica final de validación calcúlase como a media dos k resultados individuais. Neste modelo utilizamos unha versión estratificada da validación k-fold (StratifiedKfold). É unha variedade que devolve particións estratificadas, é dicir, que manteñen a porcentaxe de casos positivos e casos negativos nas particións de validación e no conxunto de datos orixinal<sup>19</sup>.

Figura 4. Validación StratifiedKfold



Fonte: <https://pub.towardsai.net/importance-of-k-fold-cross-validation-in-machine-learning-a0d76f49493e>

O conxunto de datos dividiuse en 5 particións, de forma que en cada interacción o modelo adéstrase con 4 e o rendemento calcúlase no restante grupo. Este proceso lévase a cabo cinco veces, cambiando a parte de test en cada rolda.

O obxectivo de utilizar este modelo é para asegurar que non se produzan desequilibrios entre as clases, é dicir, entre baixo peso ou non baixo peso. Sen el, podería coller un conxunto moi amplo sen baixo peso, penalizando os resultados do clasificador. Por tanto, permite que os resultados do modelo sexan máis obxectivos e fiables.

Por último, adestrouse e avalíouse o modelo para o caso de baixo peso e para baixo peso severo. Coa función `skf.split(X, y)` xérase, en cada unha das iteración, un conxunto de `train_index` cas mostradas usadas para adestrar o modelo e outro de `test_index` cas mostradas para validalo. O seguinte paso foi crear os subconxuntos correspondentes a:

---

<sup>19</sup>StratifiedKFold. (s. f.). Scikit-learn. [https://scikit-learn.org/stable/modules/generated/sklearn.model\\_selection.StratifiedKFold.html](https://scikit-learn.org/stable/modules/generated/sklearn.model_selection.StratifiedKFold.html)

`x_train_fold`: datos para adestrar

`x_test_fold`: datos para validar

`y_train_fold`: valores obxectivo no adestramento.

`y_test_fold`: valores reais para validar.

Isto é o procedemento que se repite 5 veces, para ir cambiando o conxunto de adestramento.

Con regresión loxística adestramos o modelo utilizando os conxuntos de adestramento e que posteriormente será avaliado cos conxuntos de test.

A función de `.score()` mostra a precisión das etiquetas que xerou o modelo. Na clasificación de múltiples etiquetas, esta é a precisión do grupo que estamos predicindo para coñecer que nivel de acertos e fallos tivo o modelo.<sup>20</sup> Isto almacenámolo na lista de `lst_acu_stratified` a través dun `.append`.

Os resultados deste primeiro modelo son os seguintes:

- **Precisión máxima:** 83,65%. Este é a mellor precisión que tivo o modelo nunha das 5 divisións. O 83,65% coincidiron cas etiquetas correctas.
- **Precisión mínima:** 82,39%. É o peor resultado que obtivo o modelo nunha das particións. Indica que o rendemento foi bastante similar en cada unha delas.
- **Precisión global:** representa a media das precisións para ter unha visión xeral da capacidade de clasificación do modelo.
- **Desviación típica:** mide a desviación das precisións conseguidas. O valor é de 0,0044, o que indica que o modelo é funcional e consistente nas divisións cos exemplos na base de datos. Sen embargo, estas métricas non eran suficientes para levar a cabo unha análise rigurosa do modelo. Por iso tamén se calculou a matriz de confusión, así como métricas de clasificación. Con esta estratexia puidemos detectar un problema: a matriz mostra que hai 132 verdadeiros positivos e ningún falso positivo, pero atopa 27 falsos negativos e ningún verdadeiro negativo. Por tanto, o modelo non estaba predicindo nunca unha das clases (a minoritaria). Se imprimimos `y_pred` (as predicións do modelo)

---

<sup>20</sup>

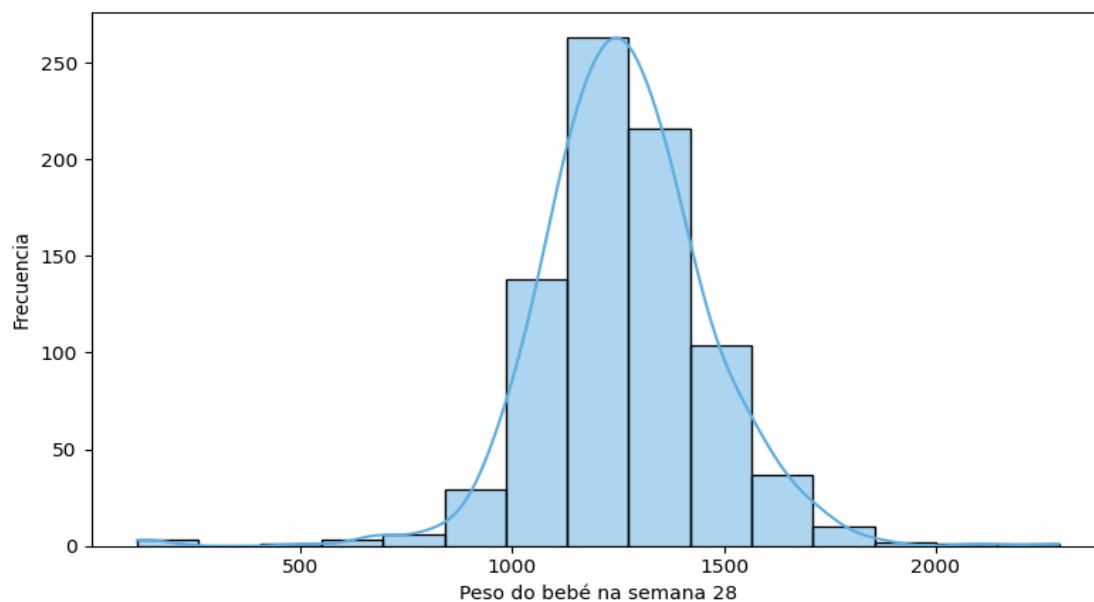
*LogisticRegression.* (s. f.). Scikit-learn. [https://scikit-learn.org/stable/modules/generated/sklearn.linear\\_model.LogisticRegression.html](https://scikit-learn.org/stable/modules/generated/sklearn.linear_model.LogisticRegression.html)

podemos observar que contén so mostras co valor 0, indicando que o modelo nunca predí a clase 1 (SGA), que é precisamente a que queremos detectar.

O modelo está aprendendo unicamente da clase maioritaria, ignorando a outra porque hai poucas observacións. A base de datos contaba con moitos máis casos de peso normal que de baixo peso ou baixo peso severo. Desta forma o modelo ía a favor da clase maioritaria e non tiña en conta a minoritaria. Esta proba evidencia a importancia de utilizar diferentes métricas de rendemento dos modelos porque neste caso, a precisión era alta (mais do 80%) e sen embargo o modelo era completamente inútil.

Na Figura 4 está representado o peso dos bebés que se atopa na semana 28. O rango adecuado nesta idade é entre 1000 e 1400 gramos<sup>21</sup>. Os datos que teñen unha frecuencia máis alta corresponden cun peso adecuado e como se ve hai moita diferenza con respecto as observacións de peso baixo. Esta representación ilustra de xeito indirecto o problema principal ao que nos enfrontamos no desenvolvemento do modelo, o desbalanceo dos datos, é dicir, a diferenza na cantidades de casos negativos e casos positivos de SGA nesta base de datos.

Figura 5. Distribución do peso fetal na semana 28



<sup>21</sup> Familia, A. (2019, 19 abril). El embarazo, semana a semana: así está el bebé en la semana 28. *Diario ABC*. [https://www.abc.es/familia/bebes/abci-embarazo-semana-semana-esta-bebe-semana-28-201904200102\\_noticia.html](https://www.abc.es/familia/bebes/abci-embarazo-semana-semana-esta-bebe-semana-28-201904200102_noticia.html)

## MODELO 2. MODELO BALANCEADO

Como solución, a propia librería ofrece a opción de utilizar o parámetro `class_weight='balance'`. Se non se indica outra cousa, en principio tódalas clases teñen o mesmo peso (igual a un). Ao engadir o modo balanceado axústanse os pesos inversamente proporcional ás frecuencias de cada clase.<sup>22</sup>

Tras implementar este importante cambio a matriz de confusión cambia. A clase 0 (o caso de non haber baixo peso) clasifica ben 87 e hai 45 falsos positivos. A clase 1 comeza a predicirse: identificáronse correctamente 15 e hai un total de 12 falsos negativos (casos que son de baixo peso, pero non se detectaron). Nas métricas detalladas, a precisión da clase 0 é bastante alta (un 88%), pero no caso da 1 é moito máis baixa, dun 25%.

Ademais o máximo de precisión correspóndese nas 5 voltas correspóndese cun 70,44% e o mínimo cun 64,15%. A desviación estándar é dun 0,02 polo que o modelo é consistente e estable.

Tamén engadimos as seguintes métricas para comprobar o funcionamento dos diferentes modelos:

Ademais destas métricas, tamén se utilizan outras que describimos a continuación:

- **Precisión:** Trátase da relación entre os verdadeiros positivos e os falsos positivos. É a capacidade do clasificador para non equivocarse etiquetando unha mostra negativa como positiva.
- **Sensibilidade:** relación entre os verdadeiros positivos e os falsos negativos, é dicir, coñecer o número de positivos que hai.

---

<sup>22</sup> *LogisticRegression*. (s. f.-d). Scikit-learn. [https://scikit-learn.org/stable/modules/generated/sklearn.linear\\_model.LogisticRegression.html](https://scikit-learn.org/stable/modules/generated/sklearn.linear_model.LogisticRegression.html)

- **Medida-F** :trátase dunha media das dúas anteriores. O obxectivo é que sexa o máis próxima a 1 posible.
- **Soporte**: número de veces que aparece cada clase nos datos reais.<sup>23</sup>

Como conclusión deste segundo modelo afirmamos que mellora porque é capaz de detectar a clase minoritaria grazas a `class_weight= 'balanced'`. É máis obxectivo porque indica se pode haber presenza de baixo peso, aínda que na clase 1 siga tendo deficiencias para a clasificación. Podemos observar que non ten un valor de precisión alto, pero polo menos é capaz de predicir as dúas clases presentes.

### MODELO 3. REDUCCIÓN DE VARIABLES PERSOAIS

Para tratar de mellorar a precisión do modelo eliminamos as variables demográficas e persoais da nai. As variables que quedan son 'HTAcronic', 'PE', 'GDM', 'DBP20s', 'PC20s', 'PA20s', 'LF20s' e estes mesmos parámetros para a semana 28.

O rendemento global máximo é de 70,63%, mentres que o mínimo é de 64,15%. A clase 1 é recoñecida e clasifica correctamente 11 casos. Sen embargo, dentro dese grupo, hai un total de 16 que non conseguiu clasificar como baixo peso, aínda que o eran.

### MODELO 4. MEDIDAS ECOGRÁFICAS SEMANA 20 E 28

Outro experimento foi analizar como se comportada o modelo se deixábamnos unicamente as medidas relacionadas co desenvolvemento do bebé nas semanas 20 e 28. Aquí o rendemento é case igual que o do caso anterior. O modelo detecta 11 casos de clase 1, con un número elevado de fallos.

### MODELO 5. MEDIDAS SEMANA 28+PE

O seguinte experimento foi o que amosou un modelo un pouco máis balanceado. Neste as variables incluídas eran preeclampsia e as medidas da semana 28 xa que teñen máis correlación

---

<sup>23</sup> *precision\_recall\_fscore\_support*. (s. f.-b). Scikit-learn. [https://scikit-learn.org/stable/modules/generated/sklearn.metrics.precision\\_recall\\_fscore\\_support.html](https://scikit-learn.org/stable/modules/generated/sklearn.metrics.precision_recall_fscore_support.html)

que as de semanas anteriores. Este modelo mellora notablemente con respecto aos anteriores. A clase 1 detéctase mellor, con 15 casos ben preditos fronte aos 11 do modelo 4-. Tamén se reducen as clasificacións mal feitas desta etiqueta, pasando de 16 a 12. Aínda que hai 39 falsos positivos, o modelo é o máis equilibrado porque se predín ambas clases. Esta selección permitiu alcanzar unha exactitude máxima do 73,13% e unha mínima do 65,67%. Ademais a desviación é a máis baixa, dun 0,0266. A continuación, mostramos a imaxe dos resultados deste modelo:

Figura 6. Resultados do modelo 5

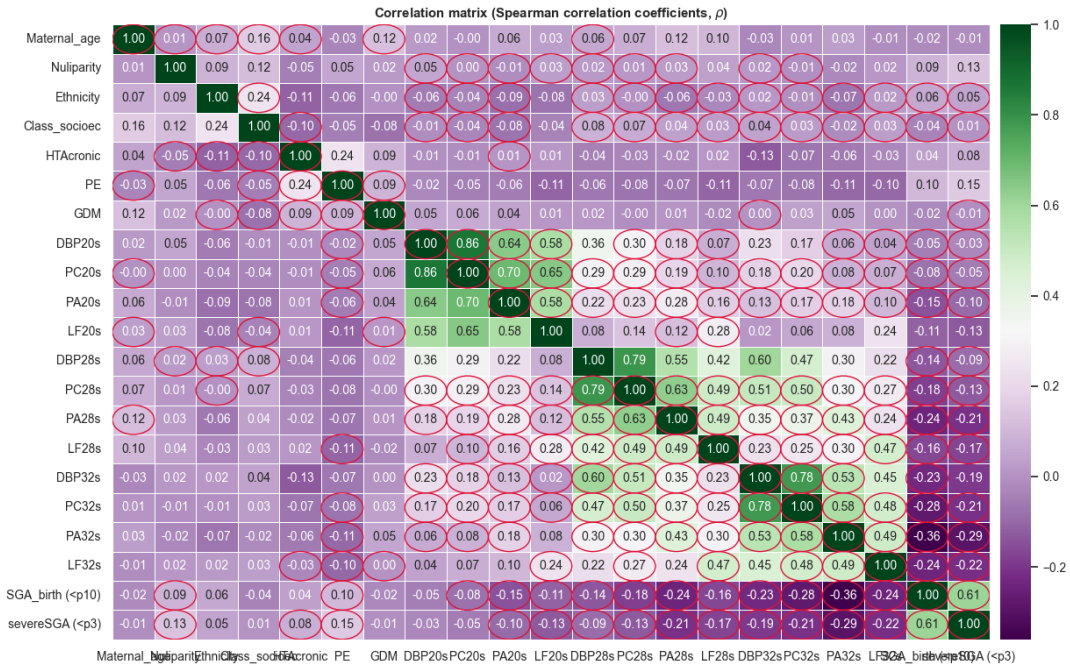


Elaboración propia

## MODELO 6. MEDIDAS SEMANA 28+32+PE

Por último, tamén engadimos as variables da semana 32 para analizar a correlación co peso baixo ou peso baixo severo. Son as que teñen valores de correlación maiores, como se ve na Figura 6.

Figura 7. Matriz de correlación de Spearman coa semana 32



Elaboración propia

A pesar de ter moita relación con esas variables e de que a exactitude se mantén, a matriz de confusión mostra peores resultados que as opcións que se presentaron anteriormente. Reduce considerablemente a clasificación correcta de “non baixo peso” pasando a 77 e os casos de “baixo peso” mantéñense. Ademais, a precisión baixa para ámbalas clases (Figura 7).

Figura 8. Resultados do modelo 6

```

PROBLEMS OUTPUT DEBUG CONSOLE TERMINAL PORTS
Exactitude: 0.6917293233082706
Matriz de confusión:
[[77 32]
 [ 9 15]]
Clasificación:
      precision    recall  f1-score   support

     0       0.90       0.71       0.79       109
     1       0.32       0.62       0.42        24

 accuracy                   0.69       133
 macro avg       0.61       0.67       0.61       133
 weighted avg    0.79       0.69       0.72       133

List of possible accuracy: [0.6865671641791045, 0.6940298507462687, 0.7313432835820896, 0.6567164179104478, 0.6917293233082706]

Maximum Accuracy That can be obtained from this model is: 73.13432835820896 %

Minimum Accuracy: 65.67164179104478 %

Overall Accuracy: 69.20772079452362 %

Standard Deviation is: 0.02658244057520504
PS C:\Users\Sandra Pena>
    
```

Elaboración propia

Como resumo do conxunto de modelos probados, axuntamos a continuación unha táboa explicativa:

Figura 9. Táboa comparativa dos modelos

MODELO	VARIABLES	PRECISIÓN MÁX/MÍN	VENTAXA PRINCIPAL	LIMITACIÓN
Básico	Variables demográficas e persoais da nai + enfermidades previas + medidas das ecografías das semanas 20 e 28.	83,65%/82,39%	Alta precisión xeral	Sesgo cara a única clase que analizaba.
Balanceado	Igual ao modelo 1 (class_weight='balanced')	70,44% / 64,15%	Considera ambas clases.	Precisión moi baixa na predición de baixo peso.
1	PE+GDM + HTAcronic+ medidas das ecografías da semana 20 e 28	70,63% / 64,15%	Simplificación das variables	Os fallos na clasificación de baixo peso son altos.
2	Medidas das ecografías da semana 20 e 28.	70,63% / 64,15%	Similar ao modelo 3.	Igual ao modelo tres
3	PE+ medidas da semana 28	71,60% / 64,43%	Mellora o balance entre as clases e a predición.	Segue presentando erros na clasificación de baixo peso.
4	PE+ medidas da semana 28 e 32	73,13% / 65,67%	Inclúe información máis recendo sobre os bebés	Peor modelo e problema cos bebés que xa naceron nesa semana.

## 3.2. Avaliación humana mediante enquisas

### 3.2.1. Preparación da enquisa

Buscamos realizar unha avaliación humana do uso da aprendizaxe automática na medicina a través de enquisas (utilizando escalas de Likert) para poder cuantificar o coñecemento das nais con respecto a situacións de nacementos de bebés con baixo peso. Esta análise permitiunos

identificar parámetros que se descoñecen durante a xestación así como o nivel de entendemento das analíticas e das probas.

Por tanto, o procedemento a seguir para a parte empírica foi:

1. Deseño experimento:

- a. Redacción dunha hipótese inicial. Redacción dunha hipótese inicial co obxectivo de contrastala ca información que se obterá das enquisas.
- b. Identificación do perfil das participantes. Definímolas en base a dous tipos de criterios: xerais e específicos. O obxectivo é garantir que a mostra sexa representativa.

Os xerais son os seguintes:

- Rango de idade ( identificar a que xeración pertencen)
- Nivel socio-cultural
- Nivel de formación académica

Os específicos son:

- Mulleres embarazadas, en calquera dos trimestres.
  - Enfermidades previas .
  - Antecedentes familiares relevantes.
- c. Definir estratexia de difusión. Algunhas das opcións que se poden levar a cabo son:
    - Mensaxes directos por medios de comunicación dixital a mulleres cun texto claro para a colaboración nun proxecto universitario, apelando á participación para coñecer o estado actual deste ámbito e a dificultade para comprendelo.
  - d. Definimos un tamaño de mostra necesario para obter datos e información que permite xerar os informes finais.

2. Elaboración da enquisa:

- a. Realización dun borrador das enquisas iniciais. Dividímolas en dúas partes. A primeira consiste na obtención de datos para trazar o perfil da persoa participante, coñecendo a situación de cada unha, así como os factores condicionantes. A outra parte, consiste na elaboración das preguntas

relacionadas cas diferentes etapas do embarazo e da comprensión dos conceptos técnicos. Inclúe preguntas relacionadas cos parámetros das analíticas, como se explican os diagnósticos de bebés que nacen con baixo peso, que dúbidas teñen con frecuencia as mulleres embarazadas se coñecen o seu significado, se serían capaces de entendela a través de comparativas sinxelas, por que datos preguntan máis...

- b. Perfeccionamento das preguntas e adaptación ao tema principal. Este apartado inclúe a redacción e reformulación das preguntas, para lograr que as participantes comprendan cada unha delas de forma sinxela.

A enquisa elaborada pódese ver no Anexo I.

3. Difundimos o enlace das enquisas en Google Forms como método de recrutamento, a través de aplicacións mensaxeiras.
4. Análise dos resultados e resumo da información obtida. Unha vez recollidas as respostas, realizamos unha análise estatística co obxectivo de identificar os principais problemas na presentación de informes médicos das ecografías, así como patróns e relacións entre a comprensión dos parámetros e o perfil das mulleres. Por último, quixemos saber en que grao están de acordo co uso da IA para a explicación dos resultados, é dicir, se prefiren o seu uso ou se, pola contra, queren mellor a explicación dun profesional.

Para estas conclusións, usamos a ferramenta de Python, calculando frecuencias e detectando relacións entre as variables.

Posteriormente, os resultados obtidos contrastámoslos coa hipótese inicial para saber se esta metodoloxía funcionou e cales son as súas limitacións principais, así como posibles melloras.

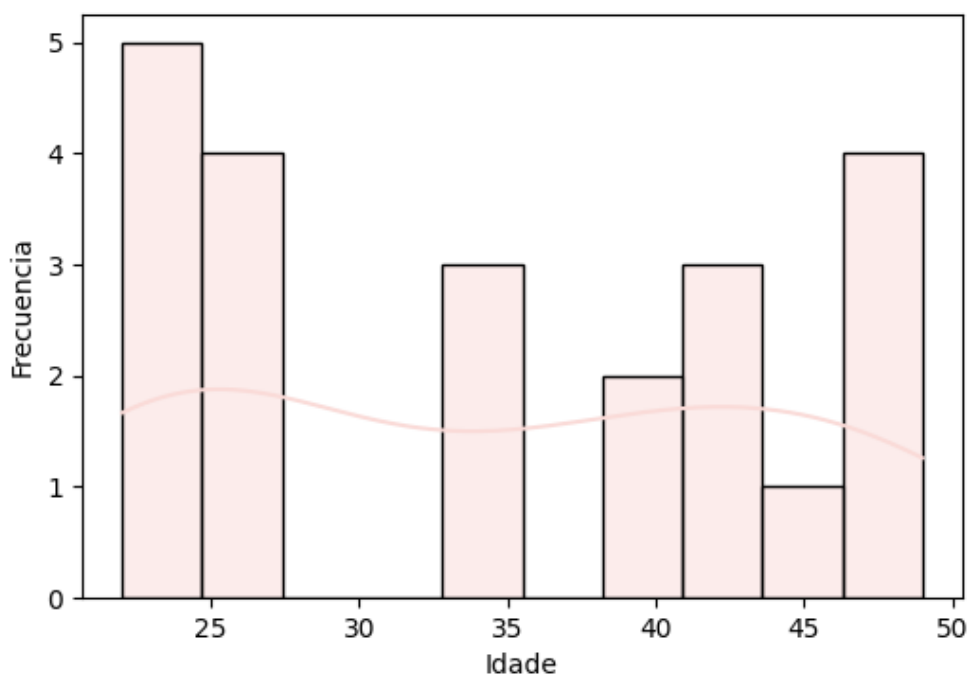
Ademais, tamén aproveitamos o clasificador para preguntar pola aceptación que teñen dos modelos de aprendizaxe automática no momento de recibir informes ou diagnósticos do embarazo. Coñecemos se están a favor dos modelos explicativos ou se, pola contra, prefiren recibir información dos médicos, sen ningún tipo de ferramenta que lles poida servir de apoio.

### 3.2.2 Análise das respostas

Na enquisa obtivemos un total de 22 respostas. Para analizar o perfil das participantes tivemos en conta as variables polas que preguntamos.

Na Figura 9 amósase a distribución da idade das participantes na mostra. É unha distribución heteroxénea, sen ningunha tendencia concreta. Sen embargo, podemos encontrar dous grupos que destacan sobre os outros. Por unha parte, as rapazas entre 22 e 27 anos, que conforman o pico máis alto da distribución. Por outra banda, arredor dos 47 aos 49 conforman a segunda frecuencia máis alta. Polo tanto, trátase dunha mostra diversa, sen centrarse nun conxunto específico, con variedade de perfís.

Figura 10. Distribución da idade das participantes



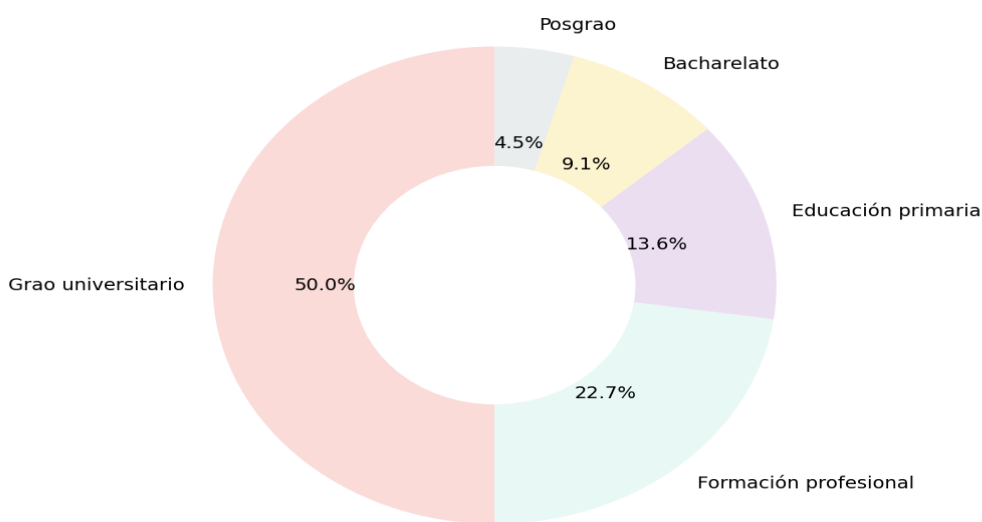
Elaboración propia

A análise da idade amosa as seguintes características:

- A media de idade das participantes é de 34,50 anos e a mediana é de 34.
- A idade mínima inscrita foi de 22 anos e a máxima de 49.
- O primeiro cuartil corresponde cos 25,25 anos, que significa que o 25% das participantes teñen esa idade ou menos, mentres que o terceiro cuartil é de 42 anos. Isto indica que o 75% están por debaixo dese valor.
- Non hai presenza de valores atípicos, pero vemos gran variedade xeracional.

En canto ao nivel educativo observamos que a maioría das participantes teñen un grao universitario, o 50% , o que supón a metade da mostra (Figura 11) . O segundo grupo máis frecuente é o da formación profesional, representando unha porcentaxe do 22,70%. Os grupos minoritarios son o de educación primaria e bacharelato que se asocian cunha educación básica ou intermedia e o de pos grao (composto por unha única persoa).

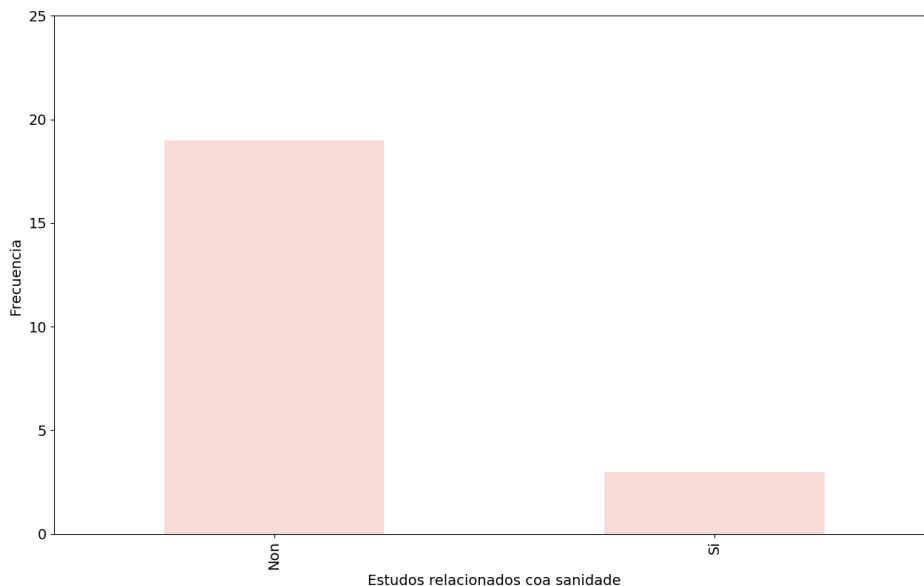
Figura 11. Nivel educativo das participantes



Elaboración propia

Se continuamos ca análise dos estudos das participantes podemos observar que a maioría non están relacionados coa sanidade. Unicamente 3 persoas das 22 enquisadas teñen coñecementos neste ámbito (Figura 12). Isto pode influír nos resultados xa que teñen máis capacidade para interpretar os parámetros que as persoas que non teñen familiaridade con estes conceptos.

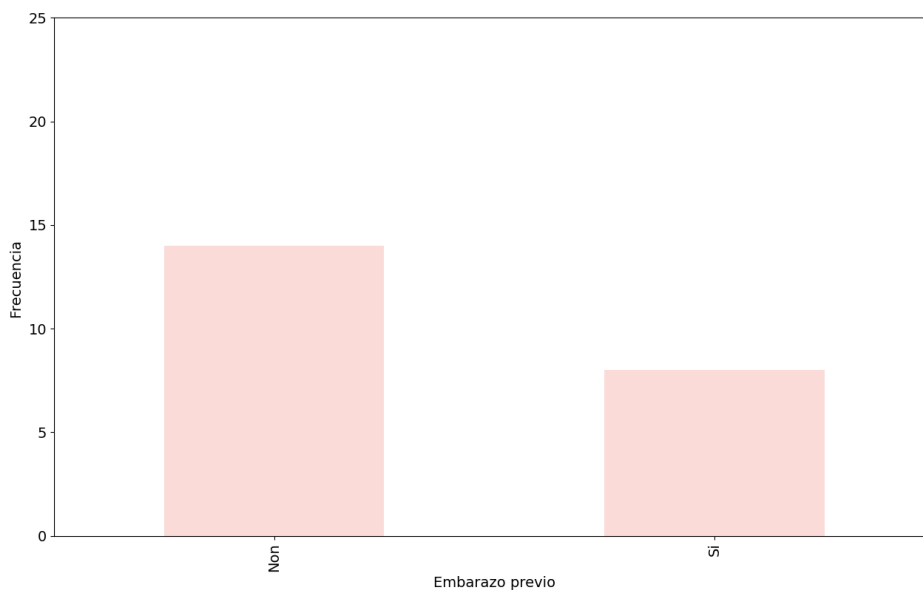
Figura 12. Participantes con estudos do ámbito sanitario



Elaboración propia

Para definir ben o perfil das enquisadas preguntámoslles se tiveran algún embarazo previo. Na seguinte imaxe podemos observar que están bastante equilibradas as dúas opcións, aínda que predomina o “non” (Figura 13). É importante ter en conta que pode deberse a que o grupo con maior número de respostas corresponde con mulleres de entre 22 e 25 anos, un grupo de rapazas novas onde ningunha delas estivo embarazada. Tamén se preguntaba se, as que tiveron un bebé, se enfrontaron a algunha complicación ao longo do embarazo. Ningunha das participantes se atopaba nunha situación, a excepción dunha que respondeu afirmativamente, pero non explicou o problema.

Figura 13. ¿As participantes tiveron algún embarazo previo?



Elaboración propia

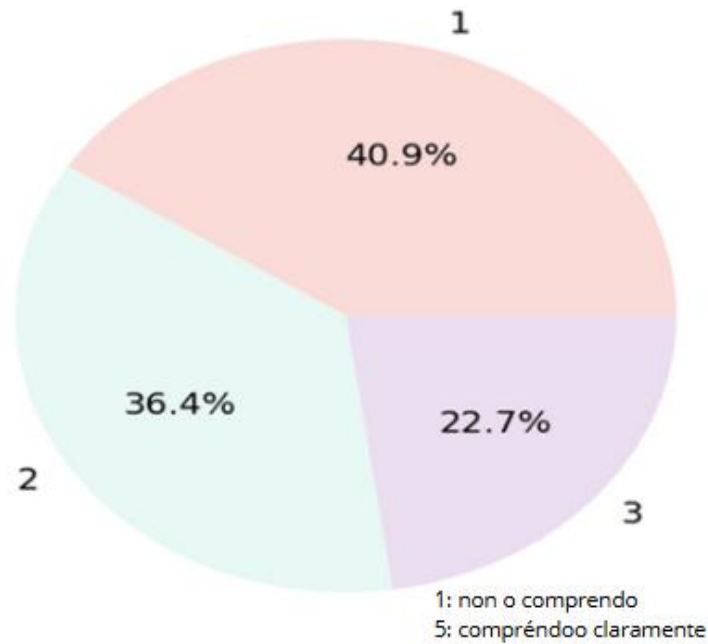
A pregunta número 6 buscaba saber en que grao as mulleres embarazadas comprenden o que significa un concepto técnico incluído nos informes que se entregan tras facer unha ecografía. O termo é traslucencia nucal que é un parámetro que se adoita utilizar para avaliar se hai risco de que o bebé teña síndrome de Down ou outras enfermidades cromosómicas <sup>24</sup>.

Tal como se observa na gráfica de tipo tarta da Figura 14, a maior parte das participantes escollen a opción 1 na escala de Likert, o que indica que non chegan a entender o significado do termo e a implicación dese dato. O valor 2 é o seguinte con maior frecuencia cun 36,4%.

---

<sup>24</sup> Examen de translucencia nucal: MedlinePlus enciclopedia médica. (s. f.).  
<https://medlineplus.gov/spanish/ency/article/007561.htm>

Figura 14. Grao de coñecemento do termo "translucencia nugal"

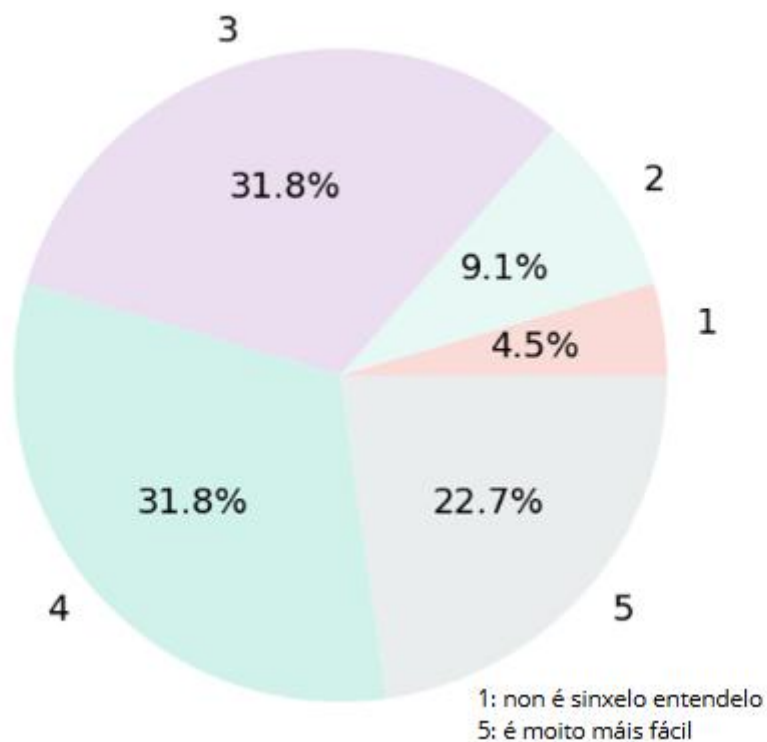


Elaboración propia

A seguinte pregunta trata sobre a facilidade de entender o concepto de "translucencia nugal" definindo que os valores normais se atopan por debaixo de 2,5 mm nun rango normal. As respostas tamén se recollían a través dunha escala de Likert. A opción máis frecuente foi o 3 que se interpreta como que non teñen unha idea clara á hora de entender este termo. O outro máis repetido é o 4 que mostra que o comprenden con maior facilidade grazas á frase. Por último, a terceira respostas con máis frecuencia é a do 5 que consideran que é moi fácil grazas ao rango.

Por tanto, as participantes manifestan maior facilidade para comprender esta frase e unicamente un 13,60% suxire dificultades.

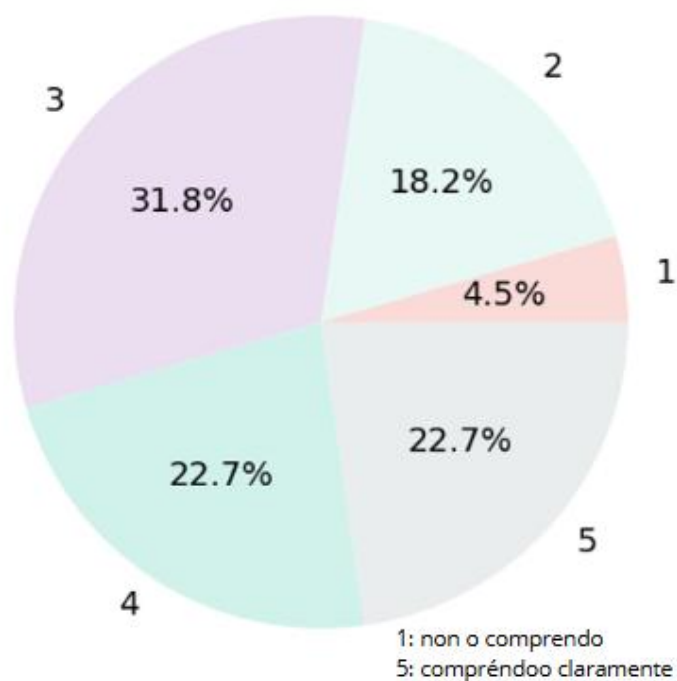
Figura 15. Capacidade de interpretar “translucencia nucal” a través de rangos numéricos



Elaboración propia

A pregunta número 8 era a seguinte: “¿Se che din que, en condicións normais, a translucencia nucal non debería ser superior ao grosor dunha uña, con que claridade entenderías este concepto?”. A metáfora resulta eficaz para un 45,4% (son as que responden o valor 4 ou 5). Axuda a facer máis sinxelos os conceptos técnicos. Non obstante, debería reforzarse con elementos numéricos xa que a pregunta anterior con rangos tivo mellores resultados. Aquí hai unha porcentaxe máis alta que responden que non o entenden (valores 1 e 2). A gráfica mostrámola a continuación:

Figura 16. Capacidade de interpretar “translucencia nucal” a través de comparacións



Elaboración propia

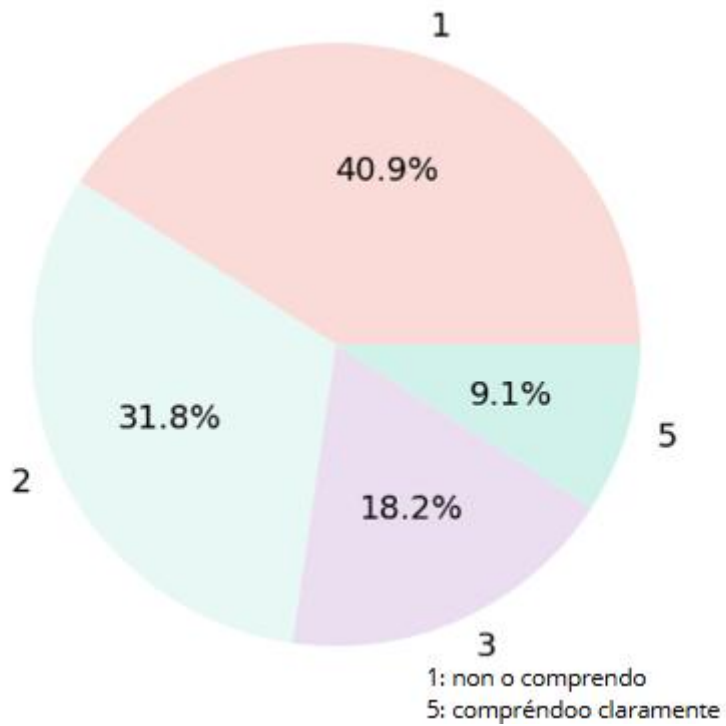
A pregunta número 10 trata sobre o diámetro biparietal que é unha medida utilizada para avaliar o crecemento do bebé durante o embarazo e que mide a distancia entre os osos parietais do cranio na parte máis ampla<sup>25</sup>.

A maioría das participantes apuntan que non teñen coñecemento deste termo e que non comprenden as implicacións que ten. Un 40,9% marcaron o valor 1, sendo case a metade das respostas. De seguido, o número con maior frecuencia é o 2 que tamén se asocia con baixa claridade (Figura 17).

---

<sup>25</sup> Diámetro biparietal: definición médica. Diccionario CUN. (s. f.-b). <https://www.cun.es>. <https://www.cun.es/diccionario-medico/terminos/diametro-biparietal>

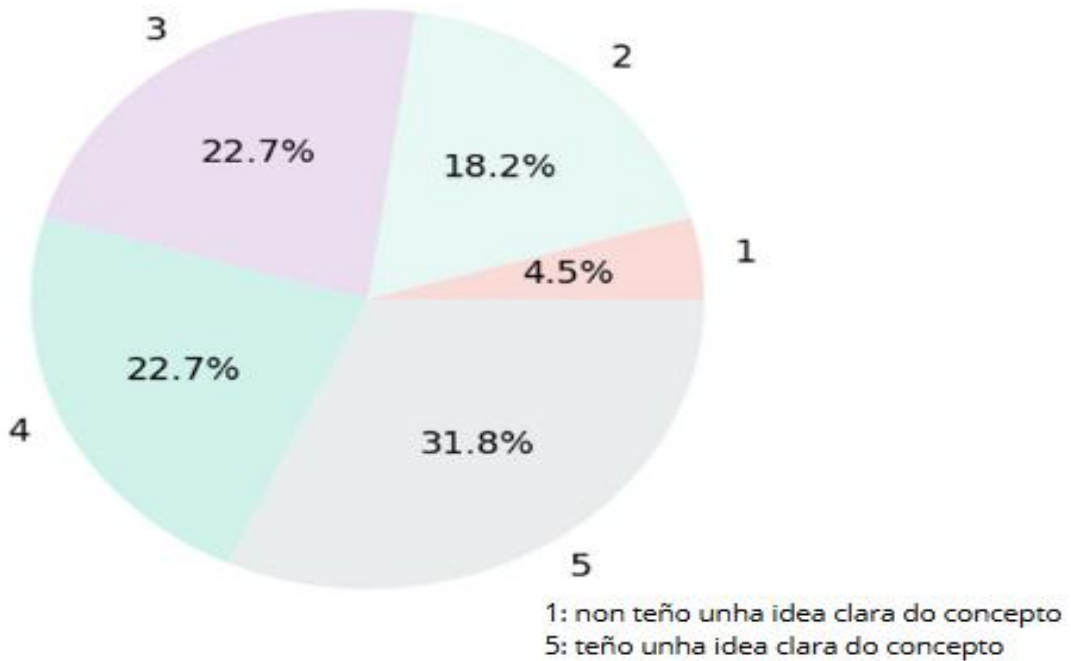
Figura 17. Grao de comprensión de “diámetro biparietal”



Elaboración propia

A seguinte gráfica móstranos que as participantes teñen maior facilidade para comprender a frase anterior grazas ao uso de rangos numéricos. Un 54,50% ten unha postura clara e afirma entender ben o concepto ( inclúense os que marcaron o 4 e o 5). Só unha pequena proporción manifesta dificultade ou dúbidas.

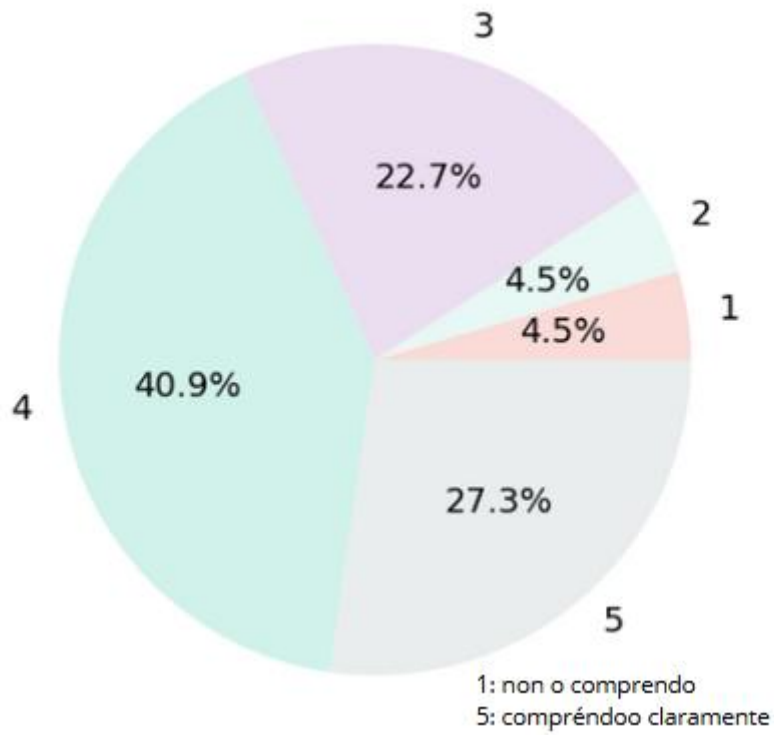
Figura 18. Grao de comprensión do diámetro biparietal a través de rangos



Elaboración propia

Os mellores resultados obtéñense facendo unha comparación ca medida do perímetro biparietal e o ancho dunha man cerrada. O valor de resposta con máis frecuencia é o 4, representado por unha porcentaxe do 40,9%. A continuación atopamos o número 5 cun 27,3%. Isto quere dicir que case todas consideran que é moi doado entender a mensaxe que transmiten os informes a través dunha comparativa, aínda que podería ter unha lixeira mellora para aquela pequena proporción que amosa dúbidas (Figura 19).

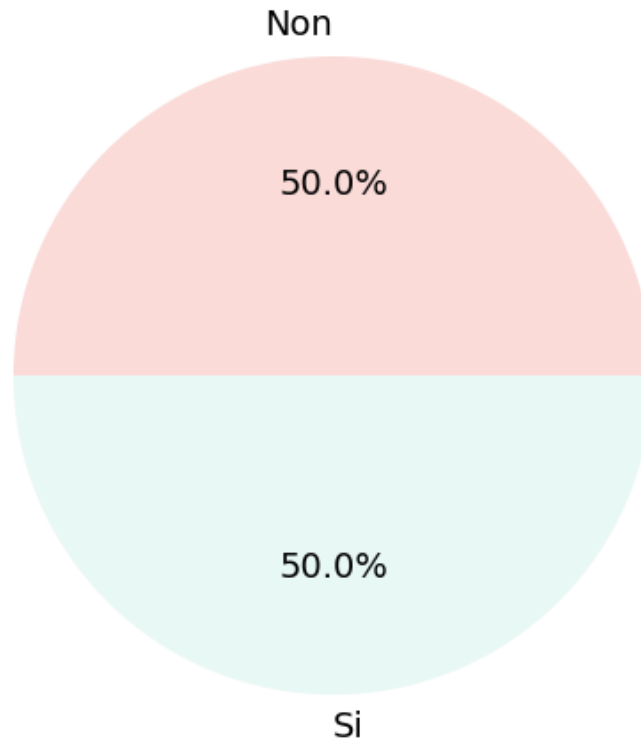
Figura 19. Grao de comprensión do diámetro biparietal a través dunha comparación



Elaboración propia

Tamén quixemos saber se as nais embarazadas se centrarían unicamente nas conclusións que traen os informes, sen analizar ningún outro parámetro. Na Figura 20 observamos que a metade estarían dispostas a ler a conclusión, sen ter en conta os valores que hai no informe.

Figura 20. Porcentaxe de participantes que se centrarían unicamente na conclusión



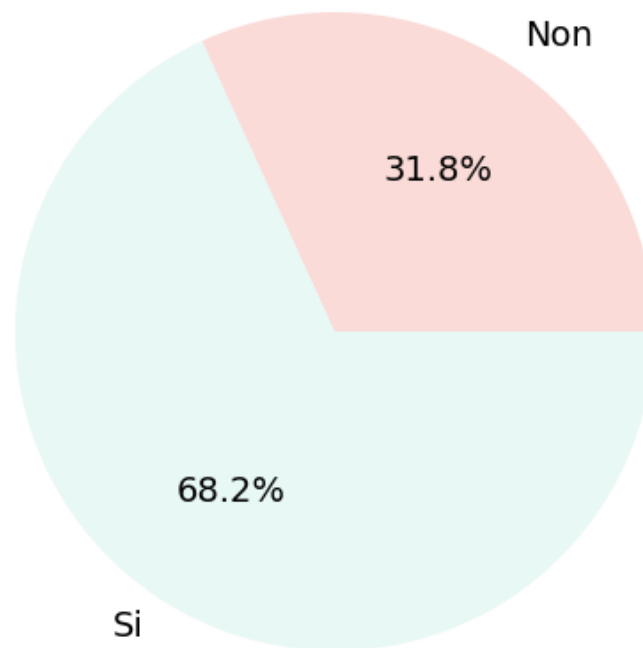
Elaboración propia

Por outra parte, analizamos se coñecen un termo máis común: os percentís. Trátase dunha medida estatística utilizada para entender o crecemento do bebé con respecto aos demais da súa idade fetal<sup>26</sup>. A pesar de que as enquisadas se moven nun rango de idade bastante amplo, un 68,2% teñen idea do que son, probablemente, porque se trata dun termo máis usado no día a día.

---

<sup>26</sup> Natalben. (2022, 4 julio). *El percentil del feto: ¿cuál es el normal para su edad gestacional?* Natalben. <https://www.natalben.com/percentiles-feto-cual-normal>

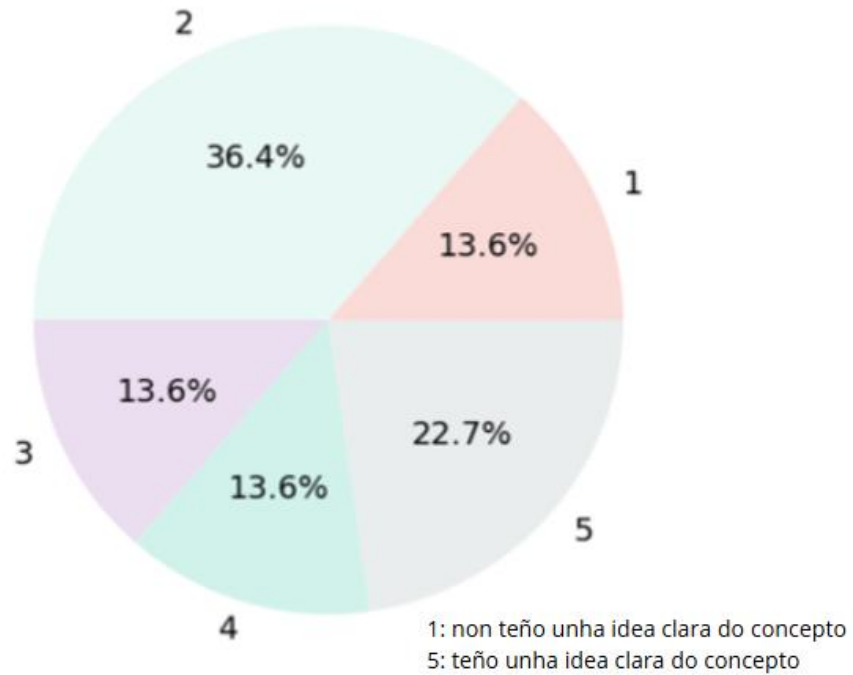
Figura 21. Porcentaxe de participantes que escoitaron falar dos percentís



Elaboración propia

A seguinte pregunta, relacionada coa anterior, pregunta se saberían explicar que significa que o bebé se atopa nun percentil baixo (p10) ou nun alto (p90). Cabe destacar que, a pesar de que a maioría escoitaron falar desta métrica, non saberían dicir que significa. O 36,4% das enquisadas responden cun valor de 2 e entendemos que teñen dificultades á hora de explicar este termo (Figura 22) .

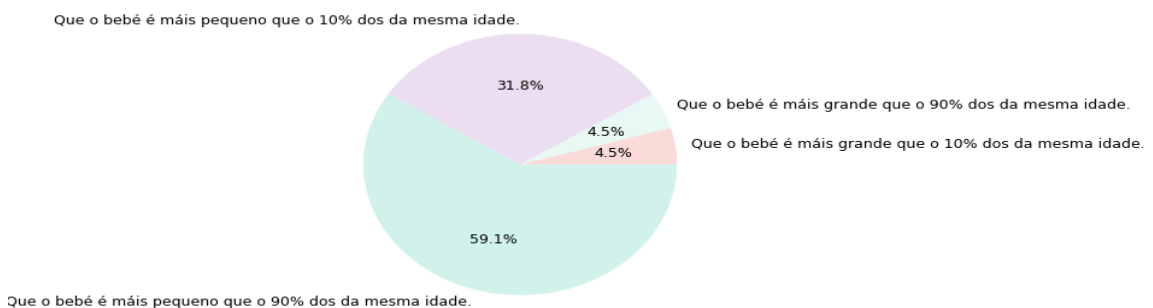
Figura 22. Porcentaxe que sabe o que significa un percentil alto ou baixo



Elaboración propia

Ademais, fixémoslles a seguinte pregunta: “Que crees que significa que se encontra no percentil 10?”. O porcentaxe total de acertos é dun 59,1% que representa unha proporción elevada con respecto ás mulleres que dicían coñecer o seu significado na pregunta anterior (Figura 23).

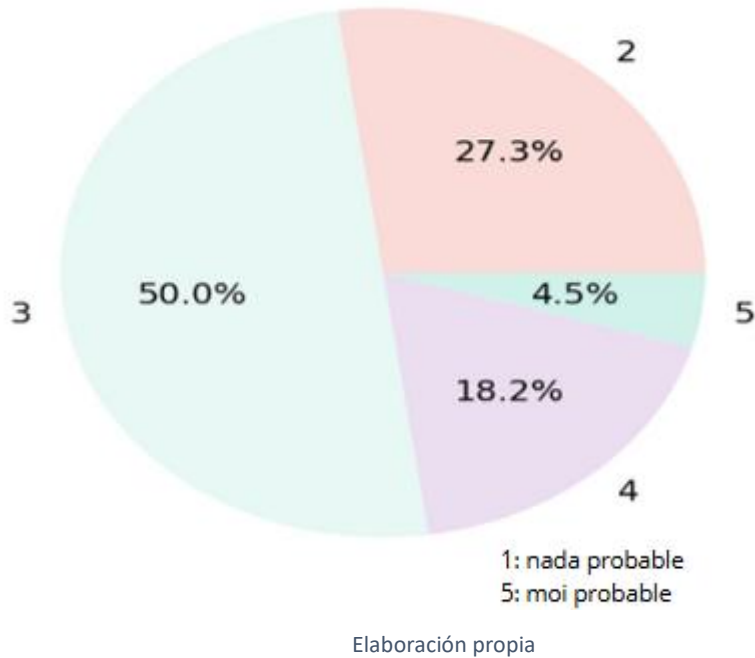
Figura 23. Respostas a pregunta de que significa que o bebé se encontra no p10



Elaboración propia

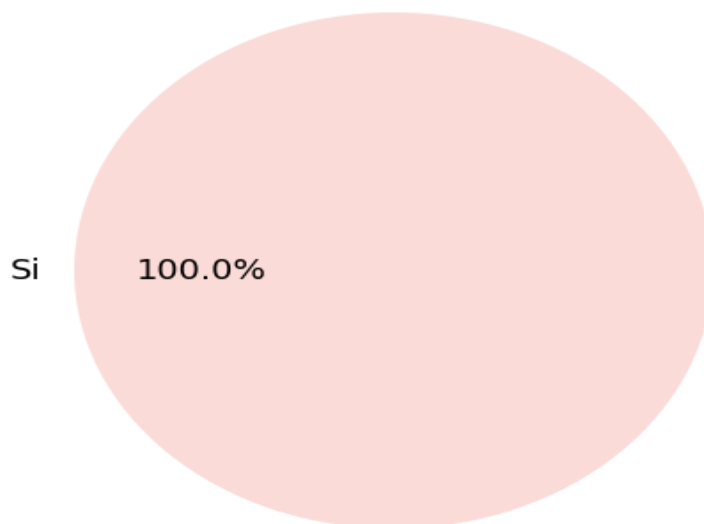
Quixemos coñecer a probabilidade de que as mulleres embarazadas buscaran o significado de todos os termos que aparecen nos informes médicos que reciben. Recollemos as respostas cunha escala e, como podemos observar, as que aseguran facelo representan unha porcentaxe moi pequena (Figura 24).

Figura 24. Posibilidade de buscar todos os parámetros dos informes



Ademais, o 100% das enquisadas aseguran que se quedarían máis tranquilas se coñeceran o significado de cada un dos termos técnicos que amosan os informes.

Figura 25. ¿quedarían máis tranquilas coñecendo tódolos parámetros?

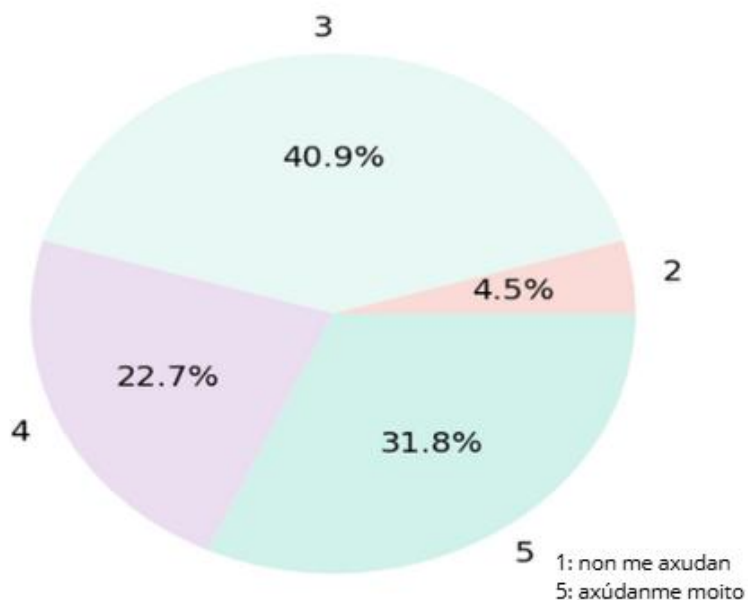


Elaboración propia

Había unha pregunta para ter suxestións de melloras nos informes médicos, pero esta non era obrigatoria. O total de respostas foi 9 e as participantes repítense: “información que comprendan os pacientes sobre que significan os parámetros”, “o rango de parámetros considerados normais” e “claridade”.

O uso de metáforas considéranlo como unha boa ferramenta para axudar a comprender os conceptos dos informes. Así o consideran un 54,50% da mostra, que marcaron o valor 4 e 5 na correspondente pregunta. Unicamente un 4,5% determinan que non é útil e ninguén marcou o 1 (Figura 26).

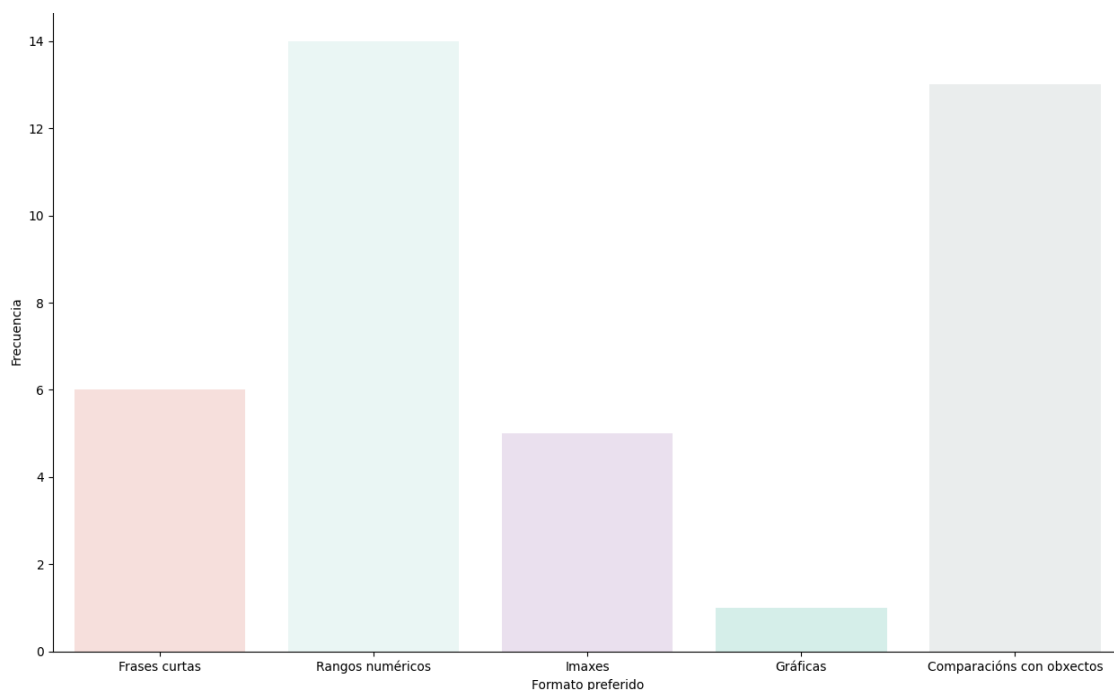
Figura 26. Como axudan os elementos cotiás a interpretación dos parámetros



Elaboración propia

As opcións preferidas para a presentación de informes son os rangos numéricos seguidos das comparacións con obxectos, é dicir, que estas dúas opcións son consideradas como óptimas para axudar a comprender conceptos que, a priori, poden resultar máis complicados por ser técnicos.

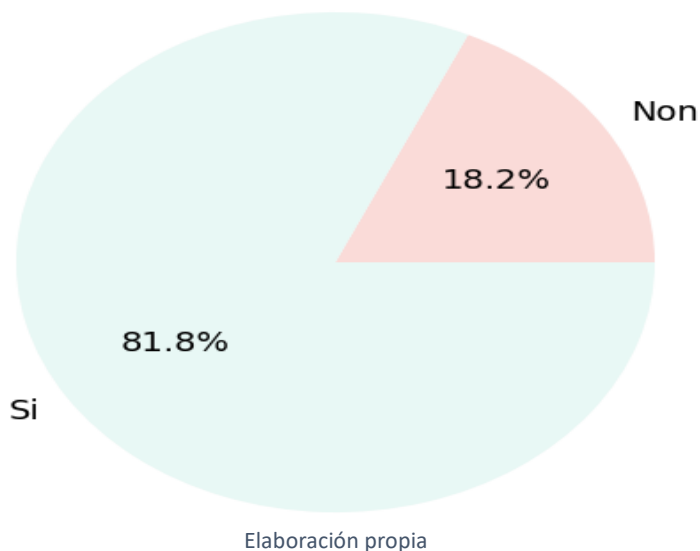
Figura 27. Preferencia de formatos de presentación de informes



Elaboración propia

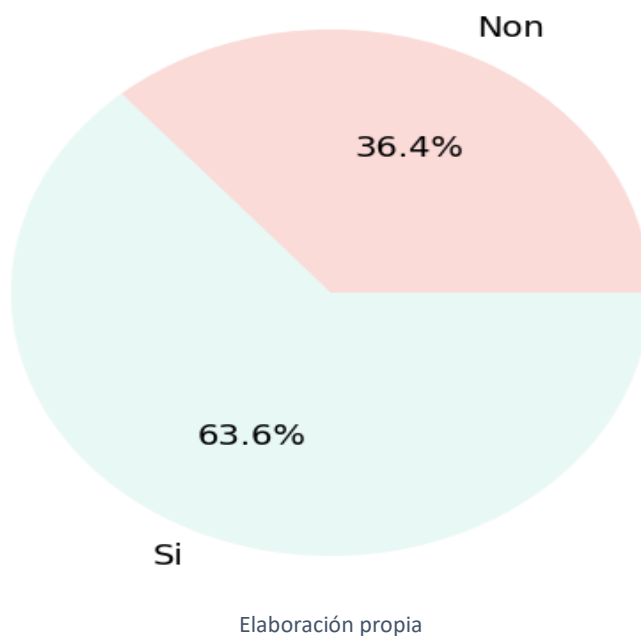
Por último, quixemos saber o grao de aceptación do uso da IA para predicións médicas. Aproveitando a análise dos nacementos con baixo peso, preguntamos se confiarían nun sistema de intelixencia artificial para interpretar os indicadores se este fora validado por médicos. A conclusión é que, se se utiliza como ferramenta de apoio xunto cos seres humanos, máis do 80% estarían a favor dela (Figura 28).

Figura 28. Aceptación dun modelo de IA con validación humana



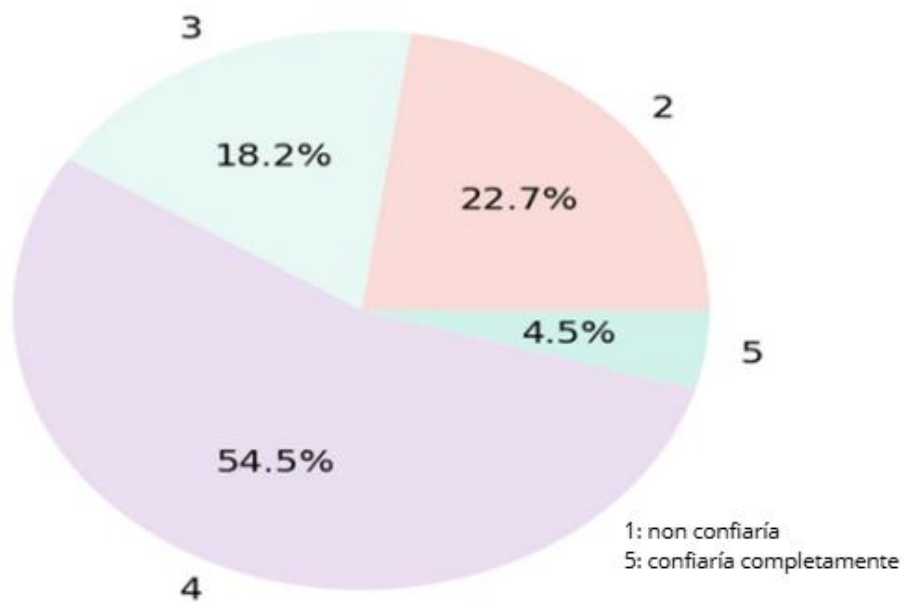
No obstante, a pesar de non rexeitar o modelo, seguen preferindo que sexan os médicos os encargados de analizar a situación de cada un e o diagnóstico. Un 63,3% optan pola opción de que os parámetros sexan explicados unicamente por médicos (Figura 29).

Figura 29. Preferencia de diagnósticos levados a cabo por médicos



Como tamén elaboramos un modelo para predicir o baixo peso, preguntamos se estarían de acordo co uso dun modelo de aprendizaxe automática cun 90% de precisión para detectar esta condición nos bebés. Máis da metade confían moito na aprendizaxe automática para facer as predicións, mentres que para un 18,2% é indiferente e un 22,70% non o usarían (Figura 30).

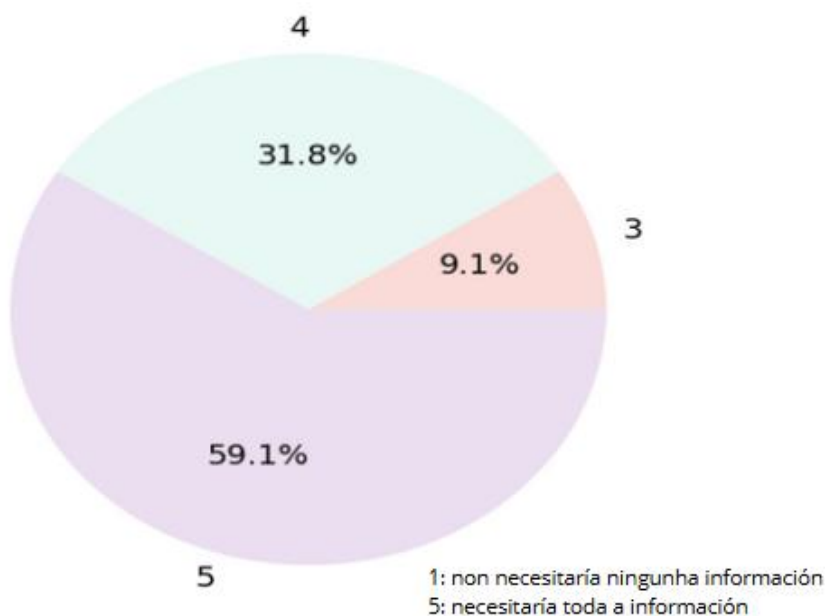
Figura 30. Grao de confianza nun modelo de IA para predicir baixo peso



Elaboración propia

Ademais, en relación co uso do modelo, un 60% da mostra necesitaría ter información de como se construíu, é dicir, cales foron as variables que se tiveron en conta á hora de entrenalo, de que poboacións e dos resultados (taxas de positivos e negativos ben clasificados ou mal clasificados). Ninguén marcou os valores 1 e 2 na escala de Likert o que indica que sen información, non estarían a favor do seu uso.

Figura 31. Necesidade de información adicional do modelo de predición de baixo peso

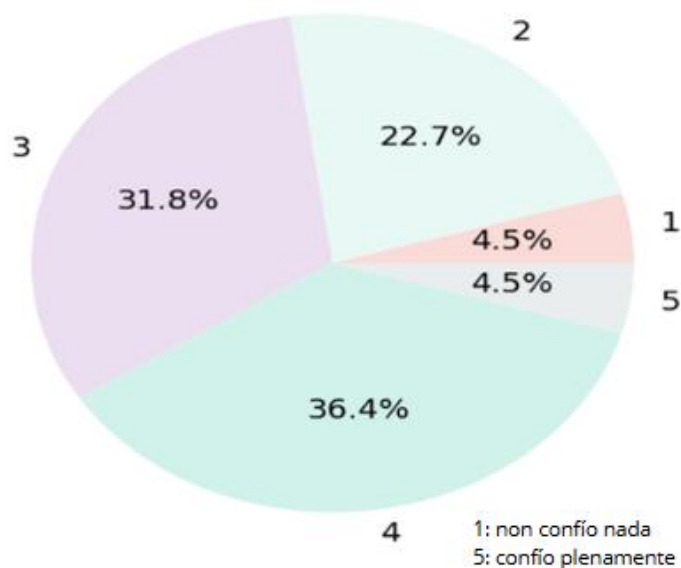


Elaboración propia

Como pregunta concluínte, analizamos o grao de aceptación de modelos de IA na medicina. Os resultados foron case ninguén está totalmente a favor ou totalmente en contra.

A actitude xeral é positiva porque un 40,90% están claramente a favor. Só un 4,5% están en contra e un 22,70% mostra certa oposición. Polo tanto, hai un grupo que desconfía do uso da IA neste campo. Tamén hai case un terzo (31,8%) que mantén unha postura intermedia e que podemos interpretar como falta de información ou que depende da situación (ferramenta de apoio médico, baseándose nos casos...).

Figura 32. Grao de aceptación da IA na medicina

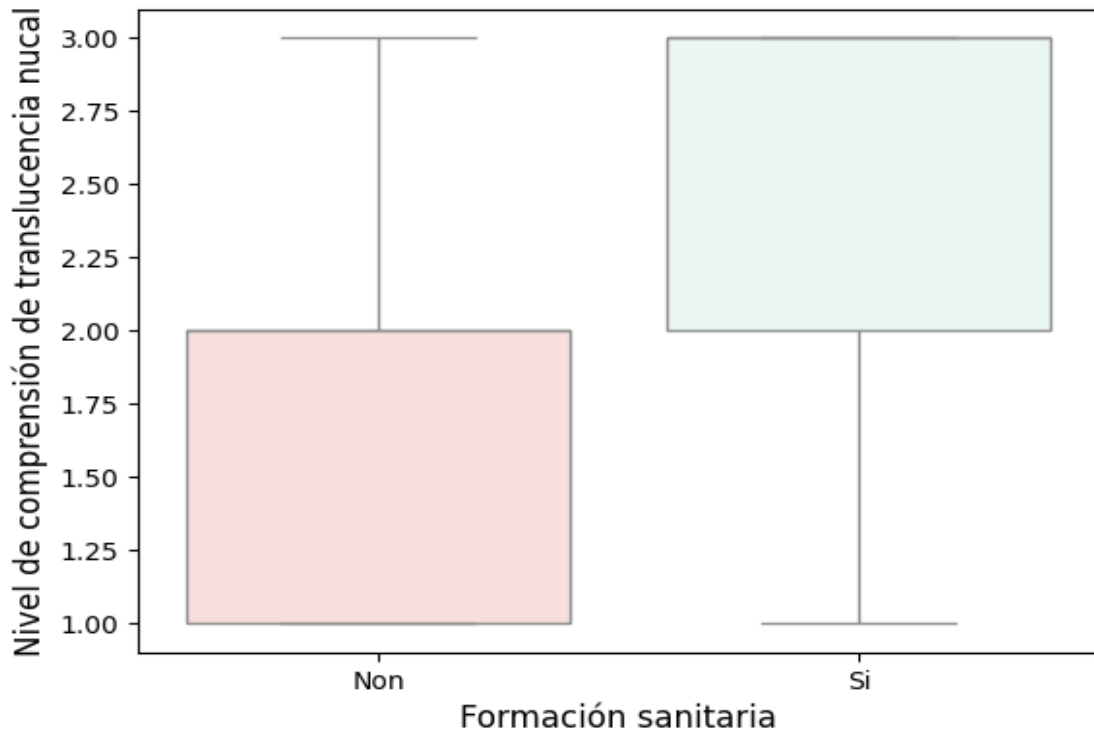


Elaboración propia

Analizamos conxuntos de variables para saber se había relación entre algunhas das que preguntamos.

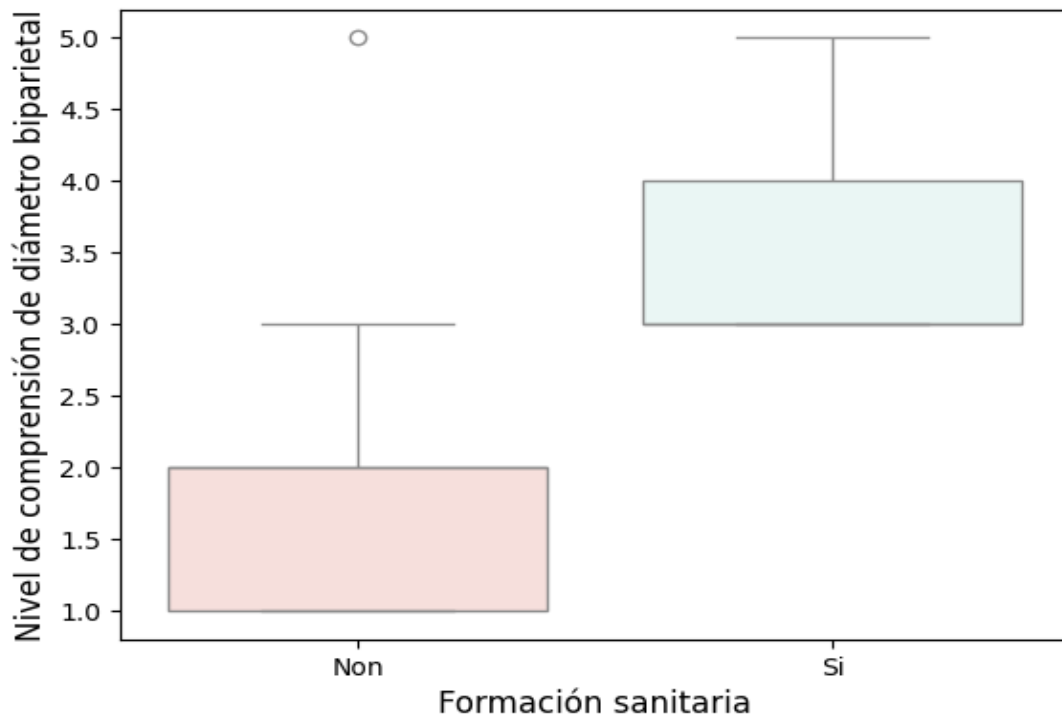
Por unha parte, xeramos un boxplot co nivel de comprensión das variables e os estudos no campo da medicina. Revelounos unha relación clara xa que aquelas que se formaron neste campo teñen maior capacidade para entender parámetros como translucencia nugal e o diámetro biparietal. Polo tanto, as persoas que xa teñen familiarización con este ámbito presentan maior facilidade para interpretar os informes. É unha maneira de ver que a presentación dos datos ten que estar adaptada a cada perfil.

Figura 33. Comprensión de translucencia nucal en relación coa formación



Elaboración propia

Figura 34. Comprensión do diámetro biparietal en relación coa formación



Elaboración propia

Tamén se fixeron probas t-test para buscar variables que estiveran relacionadas. Neste caso usamos as seguintes preguntas : “Cal é a túa idade?” e “confiarías nun sistema de intelixencia artificial para interpretar parámetros fetais se este fora validado por profesionais médicos?”. O obxectivo era determinar se existe unha diferenza estatisticamente significativa nas idades entre as persoas que din que “si” as que din que “non” no uso deste sistema de IA.

Para facelo usamos a proba estatística t-test de Student que compara dúas distribucións para saber se as medias son distintas ou se a diferenza é casualidade. O valor p é a probabilidade de que a diferenza se deba ao azar. En caso de que este valor sexa grande (maior a 0,05 normalmente), significa que é probable que se deba ao azar, mentres que se o valor é pequeno (menor que 0,05) non é casualidade<sup>27</sup>.

En Python, coa libraría `scipy.stats` importamos `ttest_ind` e traballamos como se tivera varianzas desiguais (`equal_var=False`) para identificar o valor de T e de p<sup>28</sup>. Neste caso os resultados foron os seguintes.:  $T = -4.5266$  e  $p = 0.0005$ . O valor T representa a diferenza entre as idades do grupo que dixo que si e do que dixo que non. Como se trata dun resultado negativo, indica que os que responderon afirmativamente son máis novos e ademais, ao estar lonxe de cero, a diferenza é grande. Por outra banda, o valor de p é a probabilidade de que esa diferenza sexa ou non casualidade. Como o resultado é moi inferior a 0,05, significa que non se debe ao azar senón que hai diferencias estatisticamente significativas entre os dous grupos. Podemos concluir que hai unha evidencia clara de que a idade está moi relacionada co grao de aceptación do modelo de IA, e que as mulleres novas están mais a favor da IA que as máis maiores.

Por último, buscamos algún tipo de relación entre dúas variables categóricas a través de chi cuadrado. Unha vez máis, a libraría usada foi a de `scipy.stats`<sup>29</sup>. O resultado que obtivemos destas probas foi que a aceptación da dos modelos de intelixencia artificial respaldados por

---

<sup>27</sup> *T-test con Python*. (s. f.). <https://cienciadedatos.net/documentos/pystats10-t-test-python>

<sup>28</sup> *ttest\_ind* — *SciPy v1.16.0 Manual*. (s. f.). [https://docs.scipy.org/doc/scipy/reference/generated/scipy.stats.ttest\\_ind.html](https://docs.scipy.org/doc/scipy/reference/generated/scipy.stats.ttest_ind.html)

<sup>29</sup> *chi2\_contingency* — *SciPy v1.16.0 Manual*. (s. f.). [https://docs.scipy.org/doc/scipy/reference/generated/scipy.stats.chi2\\_contingency.html](https://docs.scipy.org/doc/scipy/reference/generated/scipy.stats.chi2_contingency.html)

persoal médico e o grao de confianza xeral nesta tecnoloxía teñen unha relación estatística significativa. Os valores obtidos foron que  $\text{Chi}^2 = 16.6222$  e  $\text{P-value} = 0.0023$ . Isto implica que as persoas que queren modelos validados por médicos, tenden a ter maior grao de aceptación no uso xeral da IA en medicina. A continuación imprimimos a táboa de continxencia cos seguintes resultados:

*Figura 35. Táboa de continxencia*

<b>Aceptación IA</b>	<b>Non confía no modelo validado</b>	<b>Confía no modelo validado</b>
1	0	1
2	4	1
3	0	7
4	0	8

Elaboración propia

Con este resumo podemos concluír que as persoas que teñen maior aceptación ( que marcaron os valores 3 e 4 na enquisa), confían case completamente nos modelos de aprendizaxe automática que revisan os sanitarios xa que 15 de 15 respostas que marcan un valor igual ou superior a tres, tamén aceptan o uso da IA polos médicos.

Os códigos utilizados para facer este proxecto atópanse nos Anexos II e III.

## Conclusións, limitacións e posibles ampliacións

A investigación que levamos a cabo cumpriu con dous obxectivos: xerar un modelo de aprendizaxe automática para a predición dos nacementos con baixo peso e avaliar a través da enquisa o formato de presentación de informes e a comprensión e aceptación do uso da IA.

O modelo predictivo de SGA que utiliza as medidas ecográficas xunto coa preeclampsia foi o que ofreceu maior precisión global (71,4%) e o que presentou máis equilibrio para a clasificación da clase minoritaria (baixo peso). Isto confirmou que hai variables, como as da nai ou as medidas próximas ao parto, que achegan información e inflúen no peso que terá o bebé ao nacer. Un dos métodos que nos permitiu obter información sobre a relación entre as variables foi o cálculo da matriz de correlación cruzada de Spearman, sinalando a importancia das medidas das ecografías para determinar a que un neno teña baixo peso ao nacer.

A enquisa permitiunos saber que os rangos numéricos e as comparacións con obxectos da vida cotiá melloran a comprensión dos termos técnicos que aparecen nos informes que lle dan as futuras nais. Isto destaca a necesidade de non centrarse unicamente no valor do parámetro médico porque as pacientes non teñen coñecemento de que é nin para que serve. Elas apuntan suxestións como definir que é cada concepto, determinar cales son os valores óptimos e explicar con claridade a súa situación.

Por último, a actitude cara o uso da IA é, de forma maioritaria, positiva: un 40,9% móstrase claramente a favor. Tamén hai un 31,8% que se sitúan no punto medio da escala de Likert (valor de 3 de 5) o que pode indicarnos dúbidas, neutralidade ou que depende do contexto de aplicación. Cabe destacar que, tras a análise destes datos podemos asegurar que é necesario maior formación e claridade á hora de utilizar a IA nos contextos médicos para que os pacientes saiban como funciona na parte práctica.

En resumo, o sistema de IA é maiormente aceptado se está apoiado por profesionais médicos e acompañado dun informe que sexa comprensible. Esta é a opción que mostra maior potencia á hora de amosar os resultados ás mulleres embarazadas.

Este traballo presenta diversas limitacións que expoñemos a continuación:

- A mostra da enquisa era pequena: tivo un total de 22 participantes. Tamén hai que ter en conta que se levou a cabo nun contexto sociocultural galego e que a maioría das

participantes posen un título universitario. Isto trae consigo unha gran influencia que dificulta a extrapolación a outros lugares e a colectivos con menos nivel educativo.

- A aceptación da IA foi unha avaliación propia de cada participante. Non fixemos ningún experimento na vida real para saber como se comportaban ante a súa aplicación. Se o fixeramos, seguramente encontráramos desaxustes entre a súa reacción na realidade e o que marcan na enquisa.
- A base de datos contén unha pequena proporción de casos con baixo peso fetal. Esta desproporción de clases obrigounos a modificar o modelo e a usar diversas técnicas que permitisen coller o conxunto de datos completo, para non limitarse a un grupo con escaseza de casos de baixo peso. De todas formas, reduciu a exactitude do modelo e capacidade para predicir baixo peso, tal e como víamos na matriz de confusión.
- Excluimos variables que tiñan un elevado número de valores ausentes. Esta limitación pode levar consigo a perda de información relevante ou que podía ter patróns que non chegamos a identificar.

Por último e de cara ao futuro hai múltiples liñas de traballo para mellorar os resultados deste traballo.

Na parte empírica, pódese repetir a enquisa e ampliar a parte de percepción da IA. Un dos principais obxectivos debe ser que a mostra sexa máis grande para que desta forma se poida segmentar en grupos como a idade, os estudos ou o ámbito xeográfico. Ademais, sería interesante facer unha parte práctica (que non sexa unicamente avaliación propia) para coñecer a reacción das participantes na vida real.

Para a presentación dos parámetros a través de diferentes formatos, poderíanse crear informes que combinen elementos como gráficas, comparacións, frases ou imaxes. Así, poderíanse testar de forma práctica e identificar o grao de comprensión das usuarias nun experimento controlado como o test A/B.

En relación ao modelo de aprendizaxe automática, un obxectivo sería lograr a colaboración de máis hospitais para integrar novas bases de datos e aumentar a variabilidade dos datos. A partir disto, poderíase xerar outro modelo de aprendizaxe automática para lograr maior exactitude.

Se os resultados son bos, é importante traballar na validación en entornos clínicos reais para avaliar o impacto real na toma de decisións médicas e interpretación dos resultados prenatais.

En resumo, estas aplicacións futuras buscan consolidar a utilidade práctica do modelo á vez que se garante a aceptación das usuarias para favorecer a integración da IA neste ámbito da medicina, sempre e cando se trate dun uso regulado.

## Bibliografía

### Artigos

- Armero, W., Gray, K. J., Fields, K. G., Cole, N. M., Bates, D. W., & Kovacheva, V. P. (2023). A survey of pregnant patients' perspectives on the implementation of artificial intelligence in clinical care. *Journal of the American Medical Informatics Association*, 30(1), 46-53.
- Avila-Tomás, J. F., Mayer-Pujadas, M. A., & Quesada-Varela, V. J. (2020). La inteligencia artificial y sus aplicaciones en medicina I: introducción antecedentes a la IA y robótica. *Atención primaria*, 52(10), 778-784.
- Gambetti, A., Han, Q., Shen, H., & Soares, C. (2025). A Survey on Human-Centered Evaluation of Explainable AI Methods in Clinical Decision Support Systems. *arXiv preprint arXiv:2502.09849*.
- García-López, Andrea; Girón-Luque, Fernando & Rosselli, Diego (2023). La integración de la inteligencia artificial en la atención médica: desafíos éticos y de implementación. *Universitas Médica*, 64(3).
- Grisolía, L. P., Busselo, I. S., Zurriarán, O. M., & Murgiondo, M. I. (2008). El recién nacido de bajo peso. *Hospital Donostia. San Sebastián. UPV. Departamento de Medicina. Unidad Docente de San Sebastián. Protocolos Diagnóstico Terapéuticos de la AEP: Neonatología*, 9, 78-84.
- Lanzagorta-Ortega, D., Carrillo-Pérez, D. L., & Carrillo-Esper, R. (2022). Inteligencia artificial en medicina: presente y futuro. *Gaceta médica de México*, 158, 17-21.
- Lee, C., Willis, A., Chen, C., Sieniek, M., Uddin, A., Wong, J., ... & Gomes, R. G. (2022). Enabling faster and more reliable sonographic assessment of gestational age through machine learning. *arXiv preprint arXiv:2203.11903*.
- Modino, S. C., Martínez, M. D., González, M. B. F., Claros, N. M., Pérez, M. C. F., Aguilera, C. N., ... & Piñero, B. V. (2019). Implication of gestational diabetes treatment on maternal weight gain and low neonatal weight: a large retrospective cohort study. *Nutrición hospitalaria: Órgano oficial de la Sociedad Española de Nutrición Clínica y Metabolismo (SENPE)*, 36(6), 1261-1266.
- Osuchukwu, O. O., & Reed, D. J. (2020). Small for gestational age.

- Rivas, Marianela, González, Xiomara, & Guevara, Harold. (2016). Valores de referencia del índice de pulsatilidad de la arteria uterina durante el embarazo. *Revista de Obstetricia y Ginecología de Venezuela*, 76(4), 225-231. Recuperado en 04 de julio de 2025, de [http://ve.scielo.org/scielo.php?script=sci\\_arttext&pid=S0048-77322016000400002&lng=es&tlng=es](http://ve.scielo.org/scielo.php?script=sci_arttext&pid=S0048-77322016000400002&lng=es&tlng=es).
- Saw, S. N., Biswas, A., Mattar, C. N. Z., Lee, H. K., & Yap, C. H. (2021). Machine learning improves early prediction of small-for-gestational-age births and reveals nuchal fold thickness as unexpected predictor. *Prenatal diagnosis*, 41(4), 505–516. <https://doi.org/10.1002/pd.5903>
- Vega, M. Á., Mora, L. M. Q., & Badilla, M. V. C. (2020). Inteligencia artificial y aprendizaje automático en medicina. *Revista médica sinergia*, 5(8), 11.
- Wen, J. Y., Liu, C. F., Chung, M. T., & Tsai, Y. C. (2022). Artificial intelligence model to predict pregnancy and multiple pregnancy risk following in vitro fertilization-embryo transfer (IVF-ET). *Taiwanese Journal of Obstetrics and Gynecology*, 61(5), 837-846.

## Anexo I

### Enquisa

#### PREGUNTAS XERAIS

1. Cal é a túa idade?
2. Indica o teu nivel educativo
  - a. Educación primaria
  - b. Bacharelato
  - c. Formación profesional
  - d. Grao universitario
  - e. Posgrao
3. Algún dos teus estudos é relacionado co ámbito da sanidade?
  - a. Si
  - b. Non
4. Tiveches algún embarazo anteriormente?
5. En caso de que a resposta anterior sexa si, indica se tiveches algunha complicación

#### PREGUNTAS APLICADAS AOS INFORMES

6. A translucencia nucal é un líquido que se atopa no cuello do bebé e que se utiliza para avaliar o risco de que teña síndrome de Down ou outras enfermidades cromosómicas. Se no informe do primeiro trimestre aparece "translucencia nucal= 1,6 mm", serías capaz de entender que implica ese dato?
7. En que medida che resulta máis doado entender que "os valores de translucencia nucal inferiores a 2,5 mm atópanse nun rango normal" ?
8. Se che din que, en condicións normais, a translucencia nucal non debería ser superior ao grosor dunha uña, con que claridade entenderías este concepto?
9. Cal das seguintes opcións comprenderías con maior claridade en relación ao valor da translucencia nucal (TN) do teu bebé?
10. O diámetro biparietal é a distancia entre os ósos parietais do cráneo e úsase para avaliar o crecemento do bebé. En que medida serías capaz de entender o que implica que o diámetro biparietal é de 47 mm?
11. Se che din que o diámetro biparietal no segundo trimestre debe estar entre 43 e 53 mm, en que grado comprendes o significado?

12. Se che din que "o diámetro biparietal é o ancho da cabeza do bebé e no segundo trimestre é similar ao ancho dunha man cerrada", en que medida entendes o concepto?
13. Centraríaste unicamente nunha conclusión como " ecografía compatible coa normalidade para a idade xestacional".
14. Escoitache falar dos percentís?
15. Sabes o que significa que o teu bebé se encontre nun percentil baixo (p10) ou alto (p90)?
16. Os percentís utilízanse como medida comparativa entre o tamaño do teu bebé e o doutros da mesma idade. Que crees que significa que se encontra no percentil 10?
17. Cal é a probabilidade de que te pararas a buscar o significado de cada un dos parámetros dos informes?
18. Estarías máis tranquila se lograses entender todo o que pon no informe?
19. Que crees que sería necesario engadir nun informe fetal que ten parámetros e valores concretos, sen máis información?
20. En que medida a comparación con elementos cotiás che axuda a entender mellor os informes?
21. Que formatos preferirías para a explicación de informes deste tipo?
22. Confiarías nun sistema de intelixencia artificial para interpretar parámetros fetais se este fora validado por profesionais médicos?
23. Preferirías que o diagnóstico fose unicamente realizado por persoal médico sen intervención de IA?
24. Confiarías nun modelo de aprendizaxe automática que ten un 90 % de precisión para predicir se un bebé presenta baixo peso?
25. Para confiar nese modelo, preferirías ter algún tipo de información adicional como en que tipo de poboación foi probado (europea, americana, etc.), cal é a tasa de falsos positivos ou falsos negativos, etc.?
26. Indica o teu grao de aceptación dos modelos de IA aplicados a este ámbito da medicina

## Anexo II

### Código 1: modelo base de aprendizaxe automática

```
import pandas as pd
import numpy as np
from sklearn.model_selection import train_test_split
from sklearn.ensemble import RandomForestClassifier
from sklearn.model_selection import StratifiedKFold, KFold
from sklearn import preprocessing
from statistics import mean, stdev
```

```

from sklearn import linear_model
from sklearn.metrics import classification_report, confusion_matrix,
accuracy_score

excel = "C:/Users/Sandra
Pena/Downloads/IMPACT_gabriel_coded_20.1_final_mod_Marta.xlsx"
df=pd.read_excel(excel)

#Definir variables
variables = [
    'Maternal_age', 'Nuliparity', 'Ethnicity',
    'Class_socioec',
    'HTAcronic',
    'PE', 'GDM',
    'DBP20s', 'PC20s', 'PA20s', 'LF20s',
    'DBP28s', 'PC28s', 'PA28s', 'LF28s',
    'SGA_birth (<p10)', 'severeSGA (<p3)'
]

pd_modelo= df[variables]

si_no= ['Nuliparity', 'HTAcronic', 'GDM', 'PE', 'SGA_birth (<p10)',
'severeSGA (<p3)']
level= ['Class_socioec']
ethnicity=['Ethnicity']
print(np.unique(pd_modelo['Ethnicity']))

#Preparación de variables: categóricas a numéricas
pd_modelo= pd_modelo.replace({'yes':1, 'no':0})

for col in level:
    pd_modelo[col] = pd_modelo[col].map({'low': 0, 'medium': 1, 'high':2})

for col in ethnicity:
    pd_modelo[col] = pd_modelo[col].map({'afroam': 0, 'asian': 1, 'indi':2,
'latin':3, 'magrebi':4, 'white':5})
print(pd_modelo)

pd_modelo=pd_modelo.dropna()
print(pd_modelo.shape)
print(pd_modelo)

```

```

X = pd_modelo.drop(columns=['SGA_birth (<p10)', 'severeSGA (<p3)'])

X = X.to_numpy()
#Escalado das variables
scaler = preprocessing.MinMaxScaler()
x_scaled = scaler.fit_transform(X)

y = pd_modelo['SGA_birth (<p10)']

y=y.to_numpy()

#Aplicación do algoritmo de regresión loxística
lr = linear_model.LogisticRegression()
skf = StratifiedKFold(n_splits=5, shuffle=True, random_state=1)
lst_accu_stratified = []

for train_index, test_index in skf.split(X, y):
    x_train_fold, x_test_fold = x_scaled[train_index], x_scaled[test_index]
    y_train_fold, y_test_fold = y[train_index], y[test_index]
    lr.fit(x_train_fold, y_train_fold)

    lst_accu_stratified.append(lr.score(x_test_fold, y_test_fold))

#Imprimir valores do modelo
print('Lista de precisións:', lst_accu_stratified)
print('\nMáxima precisión:',
      max(lst_accu_stratified)*100, '%')
print('\nMenor precisión:',
      min(lst_accu_stratified)*100, '%')
print('\nPrecisión global:',
      mean(lst_accu_stratified)*100, '%')
print('\nDesviación típica', stdev(lst_accu_stratified))

```

## Código 2: modelo balanceado de aprendizaxe automática

```

import pandas as pd
import numpy as np
from sklearn.model_selection import train_test_split
from sklearn.ensemble import RandomForestClassifier
from sklearn.model_selection import StratifiedKFold, KFold
from sklearn import preprocessing
from statistics import mean, stdev
from sklearn import linear_model
from sklearn.metrics import classification_report, confusion_matrix,
accuracy_score

```

```

excel = "C:/Users/Sandra
Pena/Downloads/IMPACT_gabriel_coded_20.1_final_mod_Marta.xlsx"
df=pd.read_excel(excel)

variables = [
    'Maternal_age', 'Nuliparity', 'Ethnicity',
    'Class_socioec',
    'HTAcronic',
    'PE', 'GDM',
    'DBP20s', 'PC20s', 'PA20s', 'LF20s',
    'DBP28s', 'PC28s', 'PA28s', 'LF28s',
    'SGA_birth (<p10)', 'severeSGA (<p3)'
]

pd_modelo= df[variables]

si_no= ['Nuliparity', 'HTAcronic', 'GDM', 'PE', 'SGA_birth (<p10)',
'severeSGA (<p3)']

level= ['Class_socioec']
ethnicity=['Ethnicity']

pd_modelo= pd_modelo.replace({'yes':1, 'no':0})

""" for col in si_no_stop:
    pd_modelo[col] = pd_modelo[col].map({'yes': 1, 'no': 0, 'stop
pregnancy':2}) """

for col in level:
    pd_modelo[col] = pd_modelo[col].map({'low': 0, 'medium': 1, 'high':2})

for col in ethnicity:
    pd_modelo[col] = pd_modelo[col].map({'afroam': 0, 'asian': 1, 'indi':2,
'latin':3, 'magrebi':4, 'white':5})

pd_modelo=pd_modelo.dropna()
print(pd_modelo.shape)

X = pd_modelo.drop(columns=['SGA_birth (<p10)', 'severeSGA (<p3)'])

```

```

X = X.to_numpy()

# Normalización de datos
scaler = preprocessing.MinMaxScaler()
x_scaled = scaler.fit_transform(X)

y = pd_modelo['SGA_birth (<p10)']

y=y.to_numpy()

lr = linear_model.LogisticRegression(class_weight='balanced')      # Tendo
en conta o desbalanceo dos datos

skf = StratifiedKFold(n_splits=5, shuffle=True, random_state=1)
lst_accu_stratified = []

for train_index, test_index in skf.split(X, y):
    x_train_fold,      x_test_fold      =      x_scaled[train_index],
x_scaled[test_index]
    y_train_fold, y_test_fold = y[train_index], y[test_index]
    lr.fit(x_train_fold, y_train_fold)

    lst_accu_stratified.append(lr.score(x_test_fold, y_test_fold))
    y_pred = lr.predict(x_test_fold) # get labels instead
    print('y_test', y_test_fold)
    print('y_pred', y_pred)

    print("Exactitude:", accuracy_score(y_test_fold, y_pred))
    print("Matriz de confusión:")
    print(confusion_matrix(y_test_fold, y_pred))
    print("Clasificación:")
    print(classification_report(y_test_fold, y_pred))

print('List of possible accuracy:', lst_accu_stratified)
print('\nMaximum Accuracy That can be obtained from this model is:',
      max(lst_accu_stratified)*100, '%')
print('\nMinimum Accuracy:',
      min(lst_accu_stratified)*100, '%')
print('\nOverall Accuracy:',
      mean(lst_accu_stratified)*100, '%')
print('\nStandard Deviation is:', stdev(lst_accu_stratified))

```

## Anexo III

### Código 1: Relación entre perímetro abdominal e peso fetal estimado (semana 32)

```
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns

# Cargar os datos (cambia pola túa ruta real se o necesitas)
ruta_excel = "C:/Users/Sandra
Pena/Downloads/IMPACT_gabriel_coded_20.1_final_mod_Marta.xlsx"
df = pd.read_excel(ruta_excel)

# Columnas necesarias
x = 'PA32s'
y = 'EFW32s'

# Crear gráfica
plt.figure(figsize=(8, 5))
sns.regplot(
    x=df[x],
    y=df[y],
    scatter_kws={'color': '#fadbd8'},
    line_kws={'color': "#05c29f"},
    ci=95          #Intervalo confianza
)

plt.xlabel("Perímetro abdominal (PA32s)")
plt.ylabel("Peso fetal estimado (EFW32s)")
plt.tight_layout()
plt.show()
```

### Código 2: Matriz de correlación de Spearman (tipo para correlacións)

```
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt
import numpy as np
from scipy.stats import spearmanr
from matplotlib.patches import Ellipse
```

```

def calculate_pvalues(df, nan_policy='propagate'):
    df = df._get_numeric_data()
    dfcols = pd.DataFrame(columns=df.columns)
    pvalues = dfcols.transpose().join(dfcols, how='outer')
    for r in df.columns:
        for c in df.columns:
            pvalues[r][c] = round(spearmanr(df[r], df[c],
nan_policy=nan_policy)[1], 4)
    return pvalues

def compute_plot_correlation(data, excel_filename_corr,
excel_filename_pval, figure_filename, stat_threshold=0.001, plot=True):
    correlation = data.corr(method='spearman')
    pval_matrix = calculate_pvalues(data).astype(np.float64)
    correlation.to_excel(excel_filename_corr)
    pval_matrix.to_excel(excel_filename_pval)

    # Matriz de correlación
    fig = plt.figure(figsize=(40, 20)) # ancho e alto
    sns.set_theme(font_scale=0.9) # font size for title, and axes
labels
    ax = sns.heatmap(correlation, cmap='PRGn', annot=True,
annot_kws={"fontsize": 10}, fmt='.2f', linewidths=.5, #fontsize é tamaño
de letra, annot=true é o valor da correlación e fmt os decimais
cbar_kws={'pad': .01})

    ax.set_title(r'Correlation matrix (Spearman correlation coefficients,
 $\rho$ ), fontweight="bold")

    #Rectangulos personalizados se p>0,001
    np_pval_matrix = pval_matrix.to_numpy()
    np_corr_matrix = correlation.to_numpy()

    for i in range(np_pval_matrix.shape[0]):
        for j in range(np_pval_matrix.shape[1]):
            if np_pval_matrix[i, j] < stat_threshold:
                ax.add_patch(Ellipse((j + .5, i + .5), 1, 1, fill=False,
edgecolor='crimson', lw=1)) #destaca máis para claros e oscuros

    fig.savefig(figure_filename, bbox_inches='tight')
    if plot:
        plt.margins(x=0)
        plt.show()

excel = "C:/Users/Sandra
Pena/Downloads/IMPACT_gabriel_coded_20.1_final_mod_Marta.xlsx"
df = pd.read_excel(excel)

```

```

variables = [
    'Maternal_age', 'Nuliparity', 'Ethnicity',
    'Class_socioec',
    'HTAcronic',
    'PE', 'GDM',
    'DBP20s', 'PC20s', 'PA20s', 'LF20s',
    'DBP28s', 'PC28s', 'PA28s', 'LF28s',
    'DBP32s', 'PC32s', 'PA32s', 'LF32s',
    'SGA_birth (<p10)', 'severeSGA (<p3)'
]

pd_modelo = df[variables]

# cambia texto por números
pd_modelo = pd_modelo.replace({'yes':1, 'no':0})
level = ['Class_socioec']
for col in level:
    pd_modelo[col] = pd_modelo[col].map({'low': 0, 'medium': 1, 'high':2})
ethnicity=['Ethnicity']
for col in ethnicity:
    pd_modelo[col] = pd_modelo[col].map({'afroam': 0, 'asian': 1, 'indi':2,
    'latin':3, 'magrebi':4, 'white':5})

pd_modelo = pd_modelo.dropna()
print(pd_modelo.shape)

compute_plot_correlation(pd_modelo,
'spearman_corr.xlsx', 'spearman_corr_pval.xlsx',
                        'spearman_corr_05.png', stat_threshold=0.05,
plot=True)

```

### Código 3: Distribución do peso fetal na semana 28

```

#Importación de librerías
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt

#Ler arquivo
excel = "C:/Users/Sandra
Pena/Downloads/IMPACT_gabriel_coded_20.1_final_mod_Marta.xlsx"
df=pd.read_excel(excel)

#Lectura da variable
variables = [
    'EFW28s'

```

```

]

pd_modelo= df[variables]

#Mostrar gráfica

plt.figure(figsize=(8, 5))
sns.histplot(data=df, x='EFW28s', bins=15, kde=True, color='#5DADE2')
plt.title('Distribución do peso fetal na semana 28')
plt.xlabel('Peso do bebé na semana 28')
plt.ylabel('Frecuencia')
plt.tight_layout()
plt.show()

```

#### Código 4: Distribución da idade das participantes

```

import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns

excel = "C:/Users/Sandra Pena/Downloads/respostas.xlsx"
col = "1. Cal é a túa idade? "

df=pd.read_excel(excel)
idade= df[col]

#Cálculo das variables
n= idade.count()
media=idade.mean()
desviacion = idade.std()
min= idade.min()
max= idade.max()
q1=idade.quantile (0.25)
q2=idade.quantile (0.5)
q3= idade.quantile (0.75)

#Impresión das variables
print("O total de respostas é {}".format(n))
print("A media de idade é {:.2f}".format(media))
print("A desviación típica é {:.2f}".format(desviacion))
print("A idade mínima é {}".format(min))
print("A idade máxima é {}".format(max))
print("O primeiro cuartil é {:.2f}".format(q1))
print("O segundo cuartil ou a media é {:.2f}".format(q2))
print("O terceiro cuartil é {:.2f}".format(q3))

```

```

#Mostrar o histograma
plt.figure(figsize=(6,4))
sns.histplot(df["1. Cal é a túa idade? "], bins=10, kde=True,
color='#fadbd8')

plt.xlabel("Idade")
plt.ylabel("Frecuencia")
plt.show()

```

## Código 5: Nivel educativo das participantes (tipo dos diagramas de torta)

```

# Acceso a librarías
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns

excel = "C:/Users/Sandra Pena/Downloads/respostas.xlsx"
col = "1. Cal é a túa idade? "

#Ler arquivo
df=pd.read_excel(excel)
educacion= "2. Indica o teu nivel educativo "

#Preparación da gráfica e impresión
plt.figure(figsize=(6,6))
df[educacion].value_counts().plot(
    kind='pie',
    autopct='%1.1f%%',
    startangle=90,
    wedgeprops=dict(width=0.5),
    colors= ['#fadbd8', '#e8f8f5', '#ebdef0', '#fcf3cf', '#eaeded'],
    textprops={'fontsize': 14}
)
plt.title("Nivel educativo das participantes")
plt.ylabel('')
plt.show()

```

## Código 6: Participantes con estudos do ámbito sanitario (tipo dos diagramas de barras)

```

import pandas as pd
import matplotlib.pyplot as plt

excel = "C:/Users/Sandra Pena/Downloads/respostas.xlsx"

```

```

df=pd.read_excel(excel)
col1="3. Algún dos teus estudos é relacionado co ámbito da sanidade? "
#col2="4. Tiveche algún embarazo anteriormente? "

estudos = df[col1]
#embarazo=df[col2]

valor1= estudos.value_counts()
#valor2= embarazo.value_counts ()
valor1.plot(kind='bar', fontsize=14, color='#fadbd8')
plt.xlabel("Embarazo previo", fontsize=14)
plt.ylabel("Frecuencia ", fontsize=14)
plt.ylim(0, 25)
plt.xticks()
plt.show()

```

## Código 7: Comprensión de translucencia nucal en relación coa formación (tipo para boxplots)

```

#Importación de librarías
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt

#Ler arquivo
excel ="C:/Users/Sandra Pena/Downloads/respostas.xlsx"
df=pd.read_excel(excel)

#Resultados a través dun boxplot
sns.boxplot(
    data=df,
    x="3. Algún dos teus estudos é relacionado co ámbito da sanidade? ",
    y="6. A translucencia nucal é un líquido que se atopa no cuello do bebé e que se utiliza para avaliar o risco de que teña síndrome de Down ou outras enfermidades cromosómicas. Se no informe do primeiro trimestre aparece \"translucencia nucal= 1,6 mm\", serías capaz de entender que implica ese dato?",
    palette= ['#fadbd8', '#e8f8f5']
)
plt.xlabel("Formación sanitaria", fontsize=12)
plt.ylabel("Nivel de comprensión de translucencia nucal", fontsize=12)
plt.show()

```

## Código 8: estatística T de Student

```
#Importar librarías
from scipy.stats import ttest_ind
import pandas as pd

#Lectura de excel
excel = "C:/Users/Sandra Pena/Downloads/respostas.xlsx"
df=pd.read_excel(excel)

#Filtramos filas onde a resposta é "si" e "non", seleccionando en forma de
lista as idades que responderon "si"
si= df[df["22. Confiarías nun sistema de intelixencia artificial para
interpretar parámetros fetais se este fora validado por profesionais
médicos?"] == "Si"]["1. Cal é a túa idade? "]
non= df[df["22. Confiarías nun sistema de intelixencia artificial para
interpretar parámetros fetais se este fora validado por profesionais
médicos?"] == "Non"]["1. Cal é a túa idade? "]

#Stat para o valor T e p para a significancia
stat, p = ttest_ind(si, non, equal_var=False)
print("T = {:.4f}".format(stat))
print("p = {:.4f}".format(p))

#Imprimir resultado
if p < 0.05:
    print("O valor p é menor a 0,05. Hai unha influencia real entre a idade
e a aceptación da IA")
else:
    print("O valor p é maior a 0,05")
```

## Código 9: Chi-cuadrado

```
import pandas as pd
from scipy.stats import ttest_ind, chi2_contingency

#Datos
excel = "C:/Users/Sandra Pena/Downloads/respostas.xlsx"
df=pd.read_excel(excel)

# Cambiar nombre columnas
df = df.rename(columns={
    '22. Confiarías nun sistema de intelixencia artificial para
interpretar parámetros fetais se este fora validado por profesionais
médicos?': 'confianza_medicos',
    '26. Indica o teu grao de aceptación dos modelos de IA aplicados a
este ámbito da medicina': 'aceptacion_ia'
})
```

```
# Respostas a 0 e 1

df['confianza_medicos'] = df['confianza_medicos'].str.lower().map({'si':
1, 'non': 0})

#Tabla continxencia e impresión

tabla_chi = pd.crosstab(df['aceptacion_ia'], df['confianza_medicos'])
chi2_stat, p_chi, _, _ = chi2_contingency(tabla_chi)

print(f"Chi² statistic: {chi2_stat:.4f}")
print(f"P-value: {p_chi:.4f}")
print("Conclusión:", "Relación significativa" if p_chi < 0.05 else "Non
significativa")
print("\nTáboa")
print(tabla_chi)
```