



Traballo Fin de Mestrado

---

# Modelos de predición clínica

---

Carla Díaz Louzao

Mestrado en Técnicas Estatísticas

Curso 2015-2016



Traballo Fin de Mestrado

---

# Modelos de predicción clínica

---

Carla Díaz Louzao

Mestrado en Técnicas Estatísticas

Curso 2015-2016



## Proposta de Trabajo Fin de Mestrado

<b>Título en galego:</b> Modelos de predición clínica
<b>Título en español:</b> Modelos de predicción clínica
<b>English title:</b> Clinical prediction models
<b>Modalidade:</b> Modalidade B
<b>Autora:</b> Carla Díaz Louzao, Universidade de Santiago de Compostela
<b>Directora:</b> Carmen Cadarso Suárez, Universidade de Santiago de Compostela
<b>Tutor:</b> Francisco Gude Sampedro, Hospital Clínico Universitario da Universidade de Santiago de Compostela
<b>Breve resumo do traballo:</b> Os modelos de predición clínica estiman o risco de morte ou de presentar unha enfermidade para os pacientes de forma individualizada. Tomando como referencia dous casos clínicos (un relativo a pacientes con derrame pleural e o outro sobre pacientes sometidos a unha terapia cardíaca), considéranse os seguintes pasos no desenvolvemento de modelos de predición: formulación dos casos, especificación dos modelos, estimación, avaliación da <i>performance</i> , validación e presentación do modelo.



# Agradecementos

Debo comezar agradecendo á profesora Carmen Cadarso Suárez por aceptarme para realizar este Traballo Fin de Mestrado baixo a súa dirección. Ademais da súa axuda e apoio na realización do mesmo, este traballo permitíume experimentar o lado da aplicación práctica da estatística en servizos de consultoría, algo totalmente novo para min ata o momento.

Quero expresar tamén a miña máis sincera gratitude ó doutor Francisco Gude Sampedro, da Unidade de Epidemioloxía do Hospital Clínico Universitario de Santiago de Compostela non só polo seu apoio e a súa axuda á hora de realizar este traballo, senón por todas as actividades e proxectos nos que me incluíu maximizar e diversificar así a miña aprendizaxe durante todos estes meses.

Dou grazas tamén ó doutor Luís G. Valdés Cuadrado e á alumna de Doutoramento en Ciencias da Saúde Belén Álvarez Álvarez, quen moi amablemente me deixaron utilizar as súas bases de datos para o desenvolvemento deste proxecto, ademais de resolverme calquera dúbida con respecto ás mesmas.

Mención especial para as alumnas do programa de Doutoramento en Estatística e Investigación Operativa Jenifer Espasandín Domínguez e Ipek Güler, as cales foron de gran axuda á hora de resolver varias das dúbidas que me foron xurdindo, e, sobre todo, para os meus compañeiros do Mestrado de Técnicas Estatísticas M<sup>a</sup> Jesús Pérez Pena e Daniel Mato Regueira, que tamén estiveron na Unidade de Epidemioloxía realizando o seu Traballo Fin de Mestrado. Os tres axudámonos moito os uns ós outros, polo que este proxecto non sería o mesmo sen eles.

Por último, pero non menos importante, quero dar as grazas á miña familia e amigos por aturarme e apoiarme ó longo destes anos. Sen eles isto non sería posible.





# Índice xeral

<b>Resumo</b>	<b>IX</b>
<b>1. Introducción</b>	<b>1</b>
1.1. Estrutura do traballo . . . . .	1
<b>2. Bases motivadoras</b>	<b>3</b>
2.1. Derrame pleural . . . . .	3
2.1.1. Datos dispoñibles na Unidade de Epidemioloxía do Hospital Clínico Uni- versitario de Santiago de Compostela . . . . .	4
2.2. Terapia de resincronización cardíaca (TRC) . . . . .	8
2.2.1. Terapia de resincronización cardíaca . . . . .	9
2.2.2. Datos dispoñibles na Unidade de Epidemioloxía do Hospital Clínico Uni- versitario de Santiago de Compostela . . . . .	10
<b>3. Derrame pleural</b>	<b>13</b>
3.1. Modelo loxístico binario . . . . .	13
3.2. Curvas <i>ROC</i> . . . . .	14
3.2.1. Sensibilidade e especificidade . . . . .	15
3.2.2. A curva <i>ROC</i> . . . . .	16
3.3. Avaliación dos modelos de prognóstico . . . . .	17
3.3.1. Medidas de axuste xeral . . . . .	18
3.3.2. Medidas de capacidade de discriminación . . . . .	19
3.3.3. Avaliación da calibración . . . . .	19
3.3.4. Medidas relacionadas coa reclasificación . . . . .	20
3.3.5. Medidas relacionadas coa utilidade clínica . . . . .	22

3.4. Presentación . . . . .	23
3.5. Corrección do sobreaxuste e optimismo . . . . .	26
3.6. Resultados para os datos de derrame pleural . . . . .	27
3.7. Conclusións . . . . .	40
<b>4. Terapia de Resincronización Cardíaca</b>	<b>41</b>
4.1. Análise de supervivencia . . . . .	41
4.1.1. Censura . . . . .	41
4.1.2. Funcións básicas en análise de supervivencia . . . . .	43
4.1.3. Estimador de Kaplan-Meier . . . . .	44
4.1.4. Modelo de regresión de Cox . . . . .	45
4.1.5. Contrastes de hipóteses . . . . .	47
4.1.6. Análise de residuos . . . . .	48
4.2. Curva <i>ROC</i> dependente do tempo . . . . .	48
4.2.1. Área baixo a curva, $AUC(t)$ . . . . .	50
4.2.2. Índice de concordancia, $C$ . . . . .	50
4.3. Resultados para os datos de terapia de resincronización cardíaca . . . . .	50
4.4. Conclusións . . . . .	59
<b>5. Prácticas realizadas da Unidade de Epidemioloxía</b>	<b>61</b>
5.0.1. Valoración das prácticas . . . . .	62
<b>A. Software e código empregado</b>	<b>63</b>
A.1. Datos de derrame pleural . . . . .	63
A.2. Datos de TRC . . . . .	71
<b>Bibliografía</b>	<b>79</b>

# Resumo

## Resumo en galego

Os modelos de diagnóstico e prognóstico son moi importantes en moitos eidos, incluíndo a medicina. Este tipo de modelos, que estiman de forma individualizada o risco de morte ou de presentar unha enfermidade para os pacientes, estanse volvendo cada vez máis importantes na investigación realizada en hospitais e centros de atención primaria debido ós seus posibles usos, como son distinguir estados dunha enfermidade e clasificar a súa severidade ou avaliar a supervivencia tras un procedemento ou a diagnose dunha determinada enfermidade, xeralmente co obxectivo de tomar decisións diagnósticas ou de tratamento.

Neste traballo trátanse dous tipos de análise: un primeiro que se centra no diagnóstico dunha enfermidade en función de outras variables e un segundo que consiste nunha análise de supervivencia. Para o primeiro tipo de análise o modelo a utilizar será o modelo de regresión loxística, mentres que para o segundo o máis destacable é o modelo de regresión de Cox. En ambos casos, tamén se describirán técnicas de avaliación do modelo.

A principal motivación detrás deste traballo é a aplicación destes modelos a dous casos prácticos, un relativo ó diagnóstico de cancro e o outro relativo á supervivencia de enfermos cardíacos despois de realizarles a terapia de resincronización cardíaca ou TRC.

## English abstract

Prognostic and diagnostic models are very important in many fields, including Medicine. This kind of models, which estimate for, every individual, the death risk or the risk of having a disease, are becoming in a central issue in researching made in hospitals and primary attention centres due to their possible uses, such as distinguishing stages of a disease and classifying its seriousness or assessing the survival after a medical procedure or the diagnosis or a certain disease, usually aiming to take diagnostic or treatment decisions.

In this paperwork, two kind of analysis are treated: the first one is focused in diagnosis of a disease based on other variables and the second one consists in a survival analysis. In the first case, the model is a logistic regression model, while in the other case, the Cox regression model is the one which stands out. In both situations, some techniques for assessing the performance of the model will be described too.

The main motivation behind this work relies on the application of these models to two practical cases, one related to cancer diagnosis and the other one related to survival of heart patients after having the cardiac resynchronization therapy made.

# Capítulo 1

## Introdución

Os modelos de diagnóstico e prognóstico son moi importantes en moitos eidos, incluíndo a medicina. Este tipo de modelos, que estiman de forma individualizada o risco de morte ou de presentar unha enfermidade para os pacientes, estanse volvendo cada vez máis importantes na investigación realizada en hospitais e centros de atención primaria debido ós seus posibles usos, como son distinguir estados dunha enfermidade e clasificar a súa severidade ou avaliar o risco de presentar futuras enfermidades, xeralmente co obxectivo de tomar decisións diagnósticas ou de tratamento.

Os modelos diagnósticos tratan de identificar a existencia dunha enfermidade nun paciente ata o momento descoñecida, centrándose só no estado actual do paciente. En cambio, os modelos prognósticos tratan de ir máis aló e mirar cara ó futuro, estudando o posible curso clínico dun individuo ou a probabilidade de desenvolver unha enfermidade co paso do tempo.

Outra forma de ver os modelos de predición clínica é pensándoos como regras de predición clínica, xa que combinan certas características para tratar de predicir un diagnóstico ou un prognóstico. No caso de elaboración de diagnósticos, estes modelos son útiles para estimar a probabilidade de presentar unha enfermidade, de forma que o facultativo actúe en consecuencia con esa probabilidade, indicando un tratamento, solicitando novas probas ou considerando a posibilidade de non realizar ningunha acción futura ó respecto. Os modelos de predición tamén son moi utilizados para discernir a necesidade (ou utilidade) dun tratamento específico para un paciente, e o tipo do mesmo. Por exemplo, non se lle aplica o mesmo tratamento a unha persoa nun estadio inicial de cancro, cuxo tratamento estaría orientado principalmente a acabar coa enfermidade, que a unha persoa cun cancro moi avanzado, á que se lle aplicaríase só tratamentos paliativos.

### 1.1. Estrutura do traballo

Neste escrito trátanse dous tipos diferentes de modelos de predición clínica: o modelo loxístico, que permite estimar a probabilidade de presenza dunha certa enfermidade nun paciente, e modelos de predición en supervivencia, cos que se pretende cuantificar e describir o tempo transcorrido ata un evento de interese, xeralmente o falecemento (aínda que pode ser calquera

outro, como a reincidencia dunha enfermidade, etc.). Os gráficos realizados a modo ilustrativo da teoría foron elaborados ben con datos simulados, ben cunha submostra dalgunha das dúas bases de datos coas que se vai a traballar.

No Capítulo 2 dáse unha descrición das bases de datos motivadoras para os dous tipos de modelos (e análises) empregados, así como os obxectivos a estudar en cada caso e posibles respostas.

O Capítulo 3 céntrase na primeira base de datos, explicando primeiro en profundidade as técnicas a empregar na análise dos datos e finalizando cos resultados obtidos ó aplicar esas técnicas. Neste capítulo o modelo utilizado serán un modelo loxístico.

O Capítulo 4 está adicado á segunda base de datos, dándose primeiramente unha revisión das técnicas usadas na análise e, coma no capítulo anterior, aplicando ditas técnicas ós datos dispoñibles. Os modelos empregados neste capítulo son modelos de supervivencia.

No Capítulo 5 preséntase un breve resumo das prácticas realizadas na Unidade de Epidemioloxía do Complexo Hospitalario da Universidade de Santiago de Compostela, mostrando outras actividades realizadas ademais das dúas detalladas nos Capítulos 3 e 4.

Finalmente, o presente escrito tamén conta cun apéndice onde se trata o software utilizado e se presenta parte do código utilizado.

## Capítulo 2

# Bases motivadoras

Para a ilustración das técnicas empregadas neste traballo, empréganse dúas bases de datos: a primeira formada por pacientes que sufriron un derrame pleural e a segunda consistente en persoas con insuficiencia cardíaca sometidas á técnica TRC.

### 2.1. Derrame pleural

O corpo humano produce líquido pleural en pequenas cantidades para lubrificar as superficies da membrana que recobre os pulmóns e o interior da cavidade torácica, chamada pleura. Un derrame pleural é unha acumulación anormal e excesiva deste líquido.

Os derrames pleurais son de dous tipos:

- Un derrame pleural transudativo é causado polo líquido que se filtra cara ó espazo pleural. Isto débese a unha presión elevada nos vasos sanguíneos ou a un contido baixo de proteínas na sangue. A causa máis común é a insuficiencia cardíaca conxestiva.
- Un derrame exudativo é causado por vasos sanguíneos linfáticos bloqueados, inflamación, lesión do pulmón ou tumores.

O primeiro paso na avaliación de pacientes con derrame pleural é determinar se o derrame é transudativo ou exudativo mediante unha biopsia pleural. Despois desta proba, nun 20% dos casos aínda non se coñecen as causas do derrame. Aínda así, considérase que se o derrame pleural é transudativo non é preciso determinar a procedencia do mesmo, xa que as diferentes causas non adoitan asociarse a maiores problemas. Sen embargo, existen casos de derrames malignos que son bioquimicamente comparables ós transudativos, polo que hai un debate aberto sobre a necesidade de realizar unha citoloxía en todos os casos de derrame pleural transudativo.

Ademais, non está establecido ningún tipo de protocolo ou guía na que se indiquen os parámetros a determinar ou probas a realizar ó paciente. Unido a isto, existen tamén diferenzas

entre o tipo de unidade ou centro no que se atopa o paciente, xa que se este se atopa nunha unidade especializada con maiores recursos, poderanse estudar máis parámetros bioquímicos ca nun complexo hospitalario de menos nivel.

O obxectivo do estudo é identificar os casos de derrame pleural transudativo no que estivese xustificada a realización dunha citoloxía do líquido pleural mediante modelos de predición clínica. Para iso compáranse dous modelos: o primeiro usando só características clínicas e radiolóxicas, pensado para os centros con máis limitacións, mentres que o segundo modelo inclúe a maiores unha variable analítica (marcador tumoral) e está ideado para hospitais con máis recursos.

### 2.1.1. Datos dispoñibles na Unidade de Epidemioloxía do Hospital Clínico Universitario de Santiago de Compostela

Realizouse unha revisión retrospectiva dos rexistros médicos de todos os pacientes con derrame pleural ingresados na Unidade de Pneumoloxía do Hospital Clínico Universitario de Santiago de Compostela entre o 1 de xaneiro de 2009 e o 30 de novembro de 2015, aplicando un algoritmo diagnóstico en todos eles. En pacientes con múltiples toracocenteses, só se tomou o resultado da primeira delas.

O diagnóstico de insuficiencia cardíaca conxestiva, hidrotórax hepático, síndrome nefrótico ou síndrome do pulmón atrapado estableceuse segundo criterios coñecidos e amplamente utilizados. O fluído pleural diagnosticouse como maligno se a citoloxía ou biopsia pleural foi positiva para malignidade. Tamén se diagnosticaron outras causas para o derrame pleural mediante criterios predefinidos.

Os parámetros biomédicos empregados para distinguir entre derrame pleural exudativo e transudativo foron: cociente de proteínas (líquido pleural)/soro  $>0.5$ ; proporción de LDH (líquido pleural)/soro  $>0.6$  e LDH no líquido pleural maior ca 2/3 do límite normal superior do LDH en sangue (320 IU/L neste caso), de forma que un transudativo non cumprise ningún dos tres criterios. Aos pacientes con suposta insuficiencia cardíaca conxestiva só se lle realizou a biopsia se tiñan derrame pleural unilateral ou asimétrico bilateral, dor no peito ou febre.

Unha vez realizado todo este proceso, a base de datos final só contén datos dos pacientes con derrame pleural transudado, os cales son 281 casos.

A variable resposta:

- **mpe:** representa o feito de que o derrame pleural sexa maligno (1) ou non (0) (9.25% dos pacientes teñen derrame pleural maligno).

As variables clínicas consideradas:

- **sex:** representa o sexo do paciente (0 para mulleres e 1 para homes, cun 65.12% de homes).
- **age:** idade do paciente no momento do ingreso (media = 75.04 anos, desviación típica = 12.76 anos).



- **smoking**: variable indicadora de se un paciente é fumador (1) ou non (0). Definiuse como fumador a todo paciente que fumaba no momento do ingreso ou que deixara de fumar dentro dos 12 meses anteriores (40.93 % de fumadores).
- **time**: tempo de duración do ingreso en días (media = 21.36 días, desviación típica = 43.10 días).

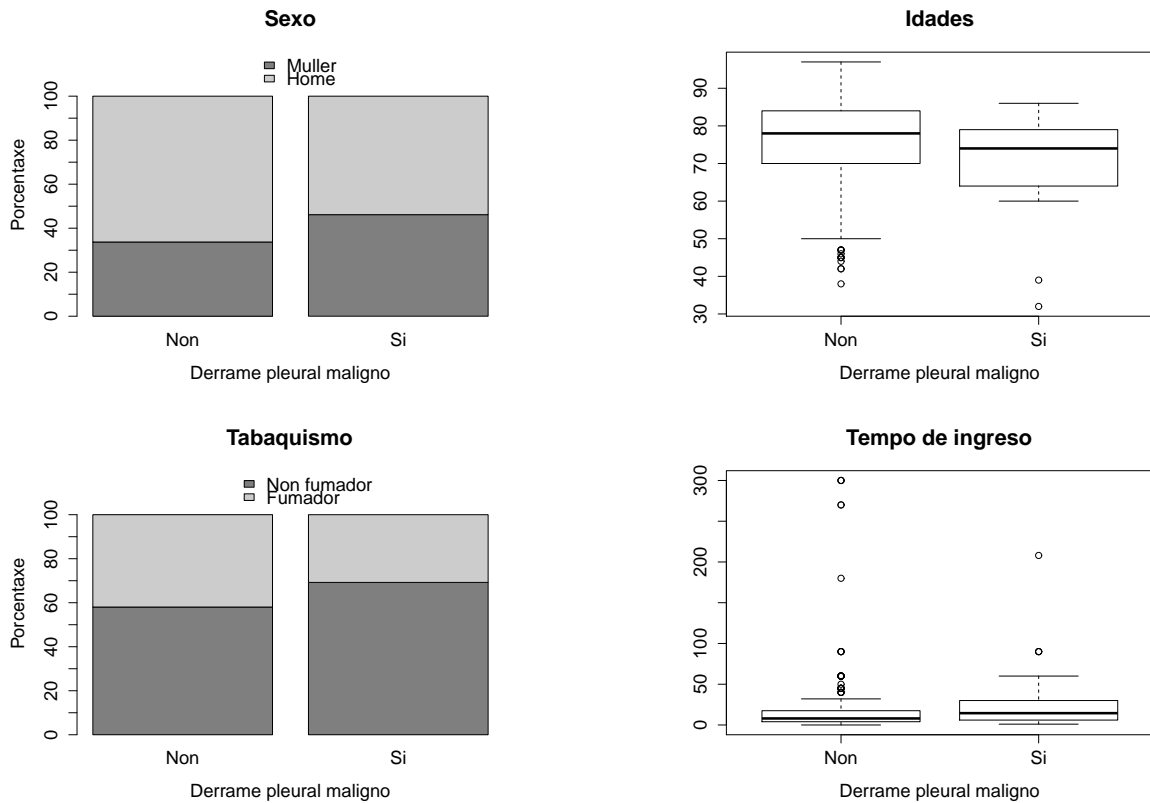


Figura 2.1.1: Representación gráfica dalgunhas das variables en función de se o derrame pleural é maligno ou non. De esquerda a dereita: *sex*, *age*, *smoking* e *time*.

- **side**: lado no que se produciu o derrame (0 se foi bilateral, 1 se foi no lado dereito e 2 se produciu no esquerdo).
- **lung.lesion**: representa a presenza de lesión pulmonar (0 se non hai, 1 no caso contrario, cun 29.89 % de pacientes con lesión pulmonar).
- **xRay**: variable categórica que marca a presenza (1) ou ausencia (0) de elementos suxestivos de malignidade tales como masas pulmonares, atelectasia pulmonar ou enfermidade dos nódulos linfáticos mediastinais nunha radiografía ou tomografía computarizada torácica (no 16.01 % dos casos se atoparon elementos suxestivos de malignidade).
- **quantity**: é a cantidade de líquido pleural codificada do seguinte xeito: 1 se o líquido ocupa ata 1/3 do pulmón ou pulmón afectados, 2 se está entre 1/3 e 2/3 e 3 se o volume

afectado é superior a 2/3 do pulmón. (o 45,91 % ocupa menos de 1/3, o 41.64 % ocupa ata 2/3 e o 12.45 % dos derrames ocupan máis de 2/3 do pulmón ou pulmón afectados).

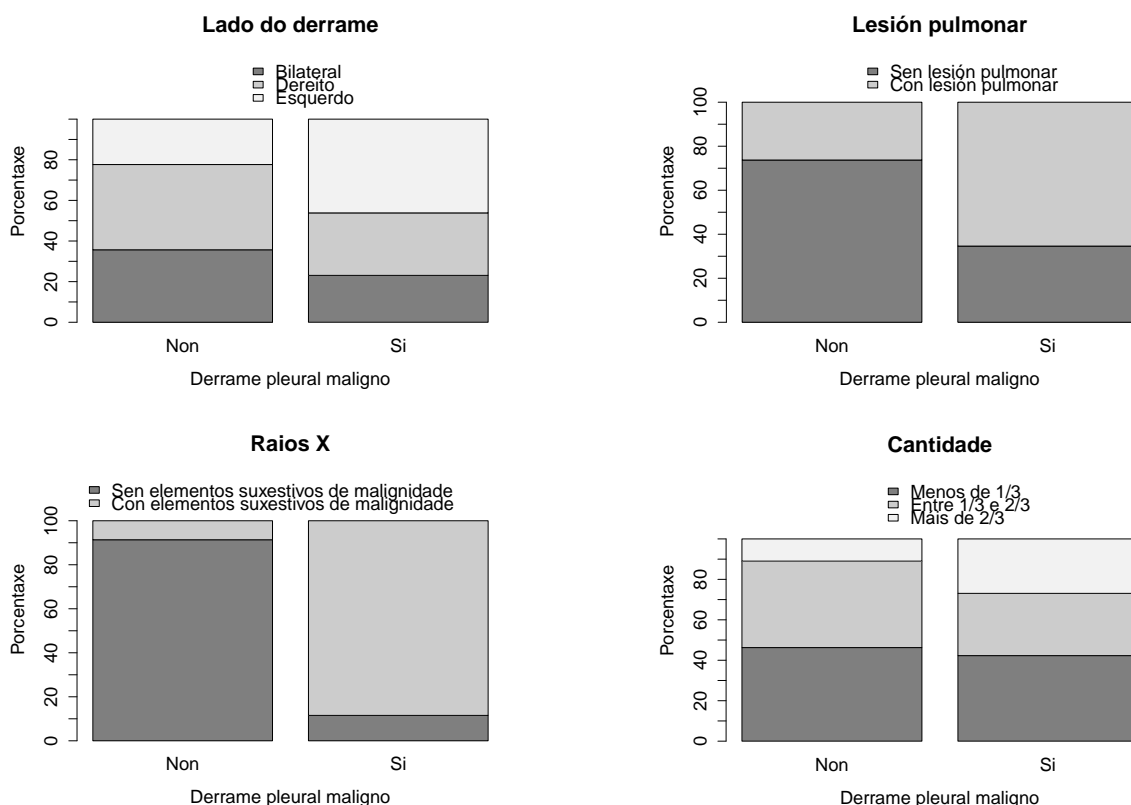


Figura 2.1.2: Representación gráfica dalgunhas das variables en función de se o derrame pleural é maligno ou non. De esquerda a dereita e de arriba a abaixo: *side*, *lung.lesion*, *xRay* e *quantity*.

- **dyspnea:** representa se o paciente ten dispnea (1) ou non a ten (0). A dispnea é un trastorno da respiración que consiste na dificultade de entrada e saída do aire dos pulmóns (o 88.61 % dos pacientes presentan dispnea).
- **pain:** variable categórica que indica a presenza (1) ou ausencia (0) de dor no peito (22.06 % de incidencia).
- **gs:** variable indicadora de se o paciente en cuestión ten síndrome xeral (1) ou non (0). Este síndrome defínese como a asociación de astenia, anorexia e perda significativa de peso (o 16.66 % dos pacientes presenta síndrome xeral).
- **fever:** variable categórica que indica se o derrame pleural do paciente cursou con febre (1) ou non (0) (17.44 % de incidencia).
- **aspect:** aspecto do fluído. Codifícase como 0 se o líquido presenta unha aparencia transparente ou limpa, mentres que 1 indica un aspecto amarelento ou sucio (o 11 % dos casos presentan líquido pleural sucio).

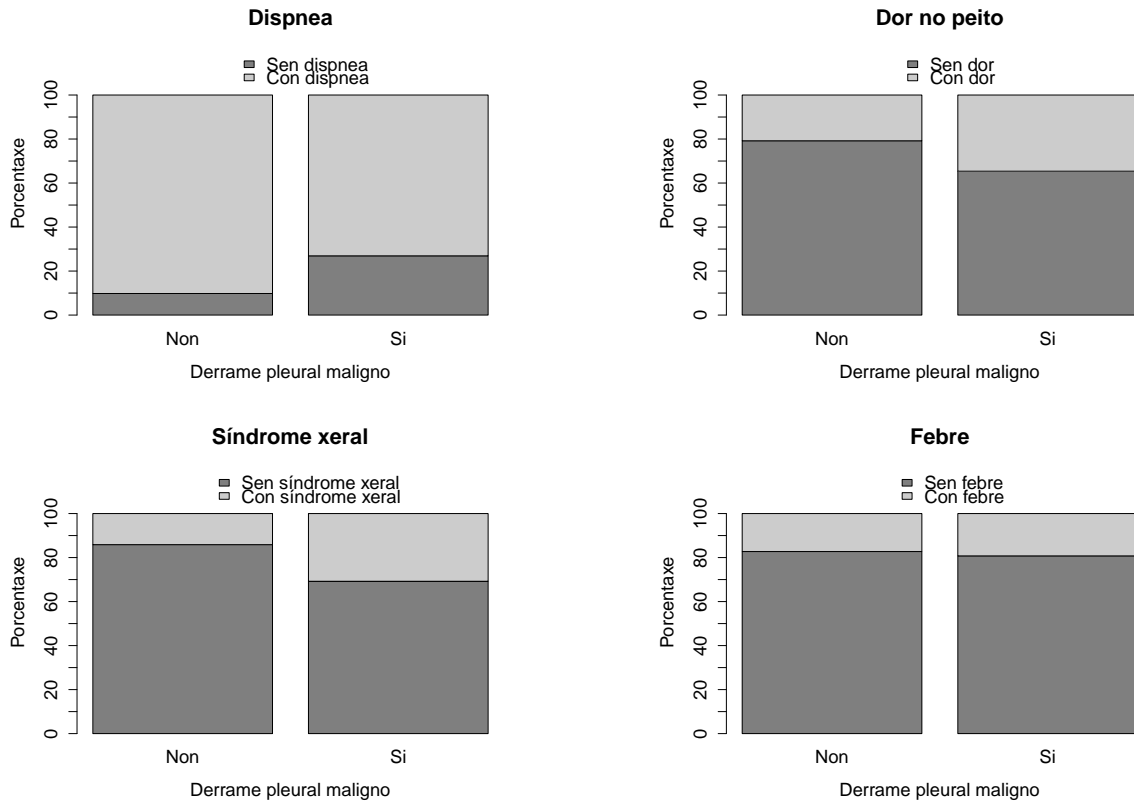


Figura 2.1.3: Representación gráfica dalgunhas das variables en función de se o derrame pleural é maligno ou non. De esquerda a dereita e de arriba a abaixo: *dyspnea*, *pain*, *gs* e *fever*.

O marcador tumoral tomado en consideración:

- **lcea:** logaritmo en base 10 dos niveis de antíxeno carcinembrionario (CEA) no líquido pleural en ng/ml (media = 0.03 ng/ml, desviación típica = 0.05 ng/ml).

Os marcadores tumorais son moi específicos pero pouco sensibles, é dicir, descartan moi ben un paciente se a causa do derrame non é cancerosa, pero detectan poucos casos nos que si o é.

As representacións gráficas das distribucións de todas estas variables en función da variable resposta (*mpe*) pódese ver nas Figuras 2.1.1, 2.1.2, 2.1.3 e 2.1.4.

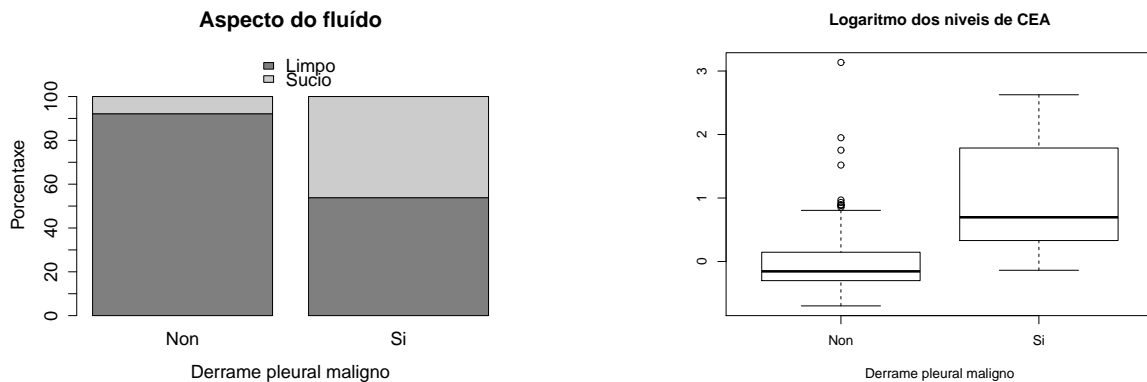


Figura 2.1.4: Representación gráfica dalgunhas das variables en función de se o derrame pleural é maligno ou non. De esquerda a dereita: *aspect* e *lcea*.

## 2.2. Terapia de resincronización cardíaca (TRC)

A insuficiencia cardíaca (IC) é unha enfermidade que ten lugar cando unha anormalidade na función cardíaca provoca que o corazón falle en impulsar sangue na taxa requirida polos tecidos para o seu metabolismo ou cando o corazón só é capaz de facelo en presenza elevada de presión de enchido. A incapacidade do corazón de bombear unha cantidade suficiente de sangue para suplir as necesidades dos tecidos orgánicos pode deberse a un defectuoso enchido cardíaco e/ou a contracción ou baleirado dificultoso.

A IC é unha das enfermidades cardiovasculares que provoca máis invalidez, maior mortalidade e maior gasto sanitario nos países desenvolvidos, cuxa prevalencia se está a incrementar nos últimos anos, chegando a acadar entre o 1 e o 2% da poboación xeral, aumentando ata o 10% nos maiores de 70 anos.

Os obxectivos do tratamento da insuficiencia cardíaca son evitar os ingresos por insuficiencia cardíaca, reducir a mortalidade e aliviar os signos e síntomas da IC. A redución dos índices de mortalidade e hospitalización por este mal reflicten a eficacia dos tratamentos.

Nos últimos anos producíronse considerables avances no tratamento farmacolóxico que reducen a mortalidade e provocan a redución nos síntomas e ingresos. Aínda así, a mortalidade por insuficiencia cardíaca segue sendo elevada.

Existe terapias non farmacolóxicas como o transplante cardíaco ou o implante de asistencias ventriculares, só consideradas en estadios avanzados de enfermidade e con acceso restrinxido.

En 2001, iniciouse unha nova etapa tras a aprobación da *U.S. Food and Drug Administration* (FDA) para o implante do primeiro dispositivo de resincronización cardíaca para o manexo da IC, tendo como obxectivos a redución da morbilidad e da mortalidade cardíacas en pacientes con esta enfermidade. Posteriormente, as Guías de Insuficiencia Cardíaca de *American College of Cardiology/American Heart Association* (ACC/AHA) do ano 2005 indicaron por primeira vez o implante de DAI (Desfibrilador Automático Implantable) e/ou TRC (Terapia de Resincronización Cardíaca) para o manexo da insuficiencia cardíaca.

A diferenciación fundamental da insuficiencia cardíaca faise a partir da fracción de execución do ventrículo esquerdo. A FE é a diferenza entre o volume diastólico final e o volume sistólico final, dividida polo volume diastólico final. Ademais de importancia prognóstica, a FE ten importancia na maioría dos ensaios clínicos. Sen embargo, é importante destacar que os valores de fracción de execución e los intervalos normais dependen das técnicas de imaxe utilizadas, o método de análise e o operador. Ademais, débese ter en conta que a gravidade dos síntomas está mal correlacionada coa función ventricular e que, pese a haber una relación clara entre a gravidade dos síntomas e a supervivencia, os pacientes con síntomas de carácter leve tamén poden presentar un risco absoluto relativamente alto de hospitalización e morte, ademais de que os síntomas poden cambiar rapidamente.

### 2.2.1. Terapia de resincronización cardíaca

Múltiples alteracións da conducción están asociadas coa presenza de disfunción ventricular, provocando á súa vez alteracións na contracción ventricular con alteracións do enchido ventricular esquerdo, redución da contractilidade e incremento da regurxitación mitral. A mediados dos anos 90, estudos de pequena magnitude demostraron o beneficio hemodinámico destes dispositivos. Actualmente, a terapia de resincronización cardíaca con estimulación biventricular demostrou ser efectiva xunto ó tratamento farmacolóxico, sendo unha estratexia de tratamento ben establecida en insuficiencia en insuficiencia cardíaca avanzada. A súa utilidade en pacientes con insuficiencia cardíaca sintomática, en clase funcional *NYHA* II-IV con disfunción ventricular esquerda e con duración do QRS prolongada ( $\geq 150$ ms), especialmente cunha morfoloxía de bloqueo de rama esquerda do Feixe de His, demostrou unha melloría na sintomatoloxía, a capacidade de exercicio, a calidade de vida e o estado funcional dos pacientes, así como a supervivencia e os parámetros de estrutura e función cardíacas a través do remodelado ventricular, tanto en pacientes con ou sen desfibrilador. Hai máis discrepancia sobre os seus efectos en pacientes en fibrilación auricular e en pacientes con bloqueo de rama dereita ou trastorno de conducción interventricular. Na Figura 2.2.1 pódese ver unha representación do corazón onde se destacan tanto o feixe de His como as dúas ramas (esquerda e dereita).

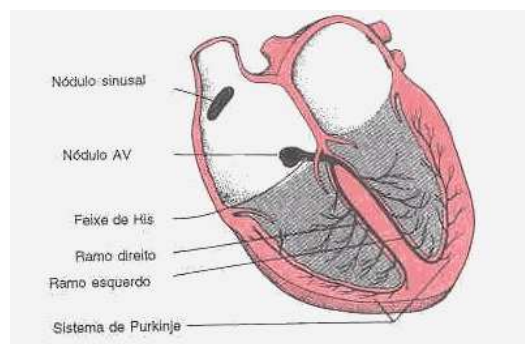


Figura 2.2.1: Representación esquemática do corazón destacándose o feixe de His e as dúas ramas do mesmo, esquerda e dereita.

Os primeiros estudos aleatorizados valoraron a resposta beneficiosa en pacientes con síntomas severos de insuficiencia cardíaca e máis recentemente en síntomas leves, demostrando

que aproximadamente un 70 % dos pacientes responden favorablemente con redución da clase funcional e ten un efecto significativo no remodelado ventricular. Pero existen aínda moitos pacientes que reciben terapia de resincronización cardíaca que non se benefician del debido á gran heteroxeneidade de resposta, polo que é preciso unha mellor selección dos criterios de implante.

En base a numerosos estudos, estableceuse como resposta a redución do volume telesistólico ventricular esquerdo en máis dun 15 % ós 6 meses e unha melloría de la fracción de exección do ventrículo esquerdo en máis dun 5 %. Ademais, recentes estudos suxeriron só beneficio da TRC en pacientes que presentan morfoloxía de Bloqueo de rama esquerda do Feixe de His.

O obxectivo principal do estudo é avaliar tanto a mortalidade como a mellora dos pacientes no seguimento a longo prazo tras o implante do dispositivo de resincronización cardíaca.

### 2.2.2. Datos dispoñibles na Unidade de Epidemioloxía do Hospital Clínico Universitario de Santiago de Compostela

Para este estudo incluíronse de forma consecutiva todos os pacientes que ingresaron no Hospital Clínico Universitario de Santiago de Compostela para implante de dispositivo de resincronización cardíaca dende agosto do 2001 ata abril de 2015. Algúns pacientes procedían doutra área sanitaria, pero tiñan como referencia o hospital anteriormente mencionado para o implante do dispositivo.

Os criterios propostos para a inclusión foron:

- Idade superior ou igual a 18 anos.
- Presencia de disfunción ventricular esquerda moderada - severa (fracción de exección menor do 40 %), tanto en ritmo sinusal coma en fibrilación auricular.
- Clase funcional *NYHA* superior ou igual a II.
- Tratamento farmacolóxico óptimo.
- Esperanza de vida superior ou igual a 1 ano no momento do implante.

En total incluíronse ó estudo 328 pacientes. Completouse o seguimento clínico no 100 % dos pacientes ata éxitus ou ata a data final de seguimento e no 99.7 % dos pacientes tívose, polo menos, un ecocardiograma durante o seguimento, habendo pacientes con ata catro ecocardiogramas.

Para a recollida das variables, revisouse de forma retrospectiva a historia clínica, tanto electrónica como en papel, dos pacientes que se someteron ó implante de TRC. Algunhas das medidas recollidas son lonxitudinais, é dicir, recolléronse diferentes medicións ó longo do tempo das mesmas. As variables a considerar son as seguintes:

- **ttot**: tempo total de seguimento dos pacientes en anos (mediana = 2.75 anos, rango intercuartílico (*IQ*) = 1.23 – 5.18 anos).

- **etiolo:** etiología da miocardiopatía: 0 para dilatada e 1 para isquémica. A miocardiopatía isquémica é un termo empregado para describir ós suxeitos cuxo corazón xa non bombea suficiente sangue ó resto do corpo debido a unha arteriopatía coronaria. A arteriopatía coronaria é un estreitamento dos vasos sanguíneos que proporcionan o sangue e o osíxeno ó corazón. En cambio, miocardiopatía dilatada é aquela na que o miocardio está debilitado e as cavidades dilatadas. Esta última é a máis usual (63.72 % de pacientes con miocardiopatía dilatada e 36.28 % con isquémica).
- **NYHA:** clasificación realizada por *New York Heart Association* para establecer a capacidade funcional de pacientes con insuficiencias cardíacas. Esta clasificación, que é a máis usada, establece catro clases (I, II, III e IV) baseándose nas limitacións na actividade física do paciente causadas polos síntomas cardíacos. As clases defínense como seguen:
  - **Clase I:** non existe limitación da actividade física. A actividade ordinaria non causa excesiva fatiga, palpitacións, dispnea ou dor anxinosa (0 % dos pacientes presentan clase I).
  - **Clase II:** lixeira limitación da actividade física. Confortables en repouso. A actividade ordinaria ocasiona fatiga, palpitacións, dispnea ou dor anxinosa (24.09 % dos pacientes presentan clase II).
  - **Clase III:** marcada limitación da actividade física. Confortables en repouso. Actividade física menor que a ordinaria provoca fatiga, palpitacións, dispnea ou dor anxinosa (70.12 % dos pacientes presentan clase III).
  - **Clase IV:** incapacidade para levar a cabo calquera actividade física sen incomodidade. Os síntomas de insuficiencia cardíaca ou de síndrome anxinoso poden estar presentes incluso en repouso. Se se realiza calquera actividade física, a incomodidade aumenta (5.79 % dos pacientes presentan clase IV).
- **tfg:** taxa de filtrado glomerular, medida en  $\text{mg} \cdot \text{ml}/\text{m}^3$ , previa ó implante. É o volume de fluído filtrado por unidade de tempo dende os capilares glomerulares renais ata o interior da cápsula de Bowman, tamén situadas nos riles. Mide a forza física e neta que produce o transporte de auga e solutos a través da membrana glomerular (mediana =  $56 \text{ mg} \cdot \text{ml}/\text{m}^3$ ,  $IQ = 44 - 76 \text{ mg} \cdot \text{ml}/\text{m}^3$ ).
- **hb:** cantidade de hemoglobina en sangue, medida en g/dL, previa ó implante (mediana = 13.3 g/dL,  $IQ = 12.0 - 14.40 \text{ g}/\text{dL}$ ).
- **ritmoimpl:** ritmo do paciente no momento do implante. Este pode ser sinusal (0), é dicir, ritmo cardíaco normal, ou fibrilación auricular ou FA (1), que se caracteriza por latidos auriculares non coordinados e desorganizados que producen un ritmo cardíaco rápido e irregular (62.5 % tiñan ritmo sinusal, mentres que o 37.5 % presentaban fibrilación auricular).
- **exitus:** indica se o paciente faleceu (1) ou seguiu vivo (0) ó final do seguimento do mesmo (o 62.20 % dos pacientes estaban vivos no momento do último seguimento).
- **BRI:** presenza (1) ou ausencia (0) de bloqueo de rama esquerda do feixe de His no momento do implante (O 60.37 % presentan bloqueo de rama esquerda).

- **sexo**: sexo dos pacientes (0 para mulleres e 1 para homes, sendo o 77.13% dos pacientes varóns).
- **edad**: idade en anos dos pacientes no momento do implante (mediana = 72 anos,  $IQ = 64 - 77$  anos).
- **fe0**: fracción de exección previa ó implante. Os valores normais son entre 80 e 90 (mediana = 28,  $IQ = 21 - 32$ ).
- **time**: o número de consulta ou visita. Como xa se dixo, o 99.7% teñen polo menos unha visita, podendo chegar a ter ata 4 visitas.
- **fe**: fracción de exección, medida en cada consulta posterior ó implante que tivo o paciente. Na parte esquerda da Figura 2.2.2 pódense ver os perfís evolutivos da fracción de exección de todos os pacientes, mentres que na parte dereita se ten unha estimación non paramétrica da evolución da fracción de exección ó longo do tempo, con intervalos de confianza ó 95%.
- **deltafe**: porcentaxe de variación da fracción de exección, medida en cada consulta posterior ó implante que tivo o paciente.

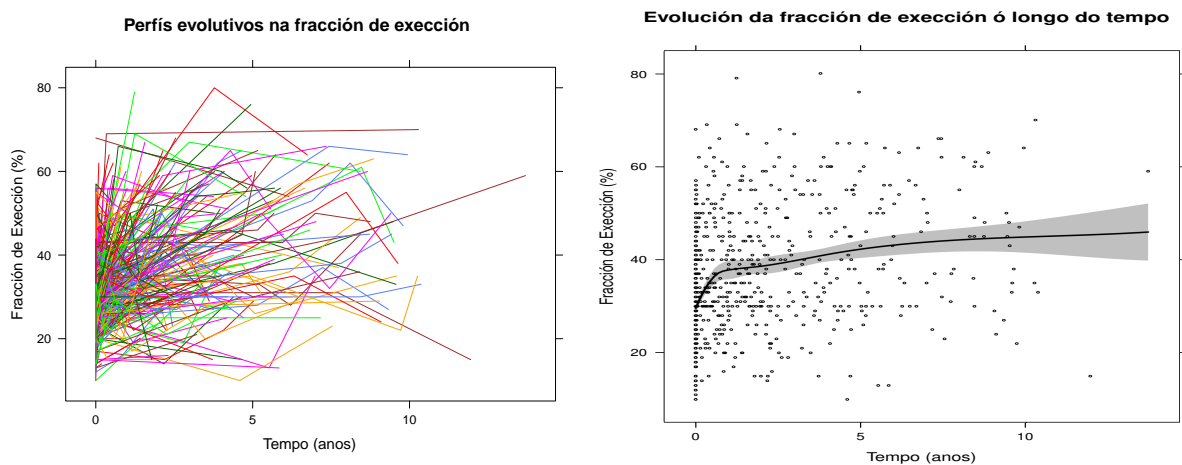


Figura 2.2.2: Esquerda: perfís evolutivos da fracción de exección. Dereita: Evolución da fracción de exección ó longo do tempo, con bandas de confianza ó 95%.



## Capítulo 3

# Derrame pleural

Neste capítulo vaise estudar a base de datos relativa aos derrames pleurais transudados, a cal foi explicada no Capítulo 2. Como xa se dixo anteriormente, a variable resposta é binaria, polo que a o modelo a utilizar será un modelo de regresión loxística.

Os modelos de regresión loxística son amplamente utilizados en medicina debido a que son moi útiles para os casos nos que a variable resposta consiste nun marcador de presenza ou ausencia dunha enfermidade (no caso deste escrito, presenza ou non dunha causa maligna do derrame pleural), os cales son moi frecuentes neste eido. Para este tipo de variables non se pode usar un modelo de regresión linear múltiple, só apto para respostas continuas, polo que é necesario a implementación dun novo modelo.

### 3.1. Modelo loxístico binario

Considérese unha variable resposta  $Y$  que só pode tomar dous valores, codificados como 0 e 1. Dirase que  $Y$  é unha variable binaria ou dicotómica, da cal se quere estudar a súa relación con outras variables explicativas  $X$  mediante unha análise de regresión. Por ser a variable resposta binaria, a súa distribución é Bernoulli e a súa media unha probabilidade de éxito.

O obxectivo é construír un modelo para

$$\pi(x) = \mathbb{P}(Y = 1/X = x)$$

é dicir, para a probabilidade de éxito condicionada a cada valor da variable explicativa. Debido á natureza da variable resposta, se se intentase axustar un modelo de regresión linear directamente, estaríanse incumprindo tres das catro hipóteses do modelo (linearidade, homocedasticidade e normalidade). Por tanto, se se desexa considerar un modelo linear, pódese aplicar previamente a  $\pi(x)$  unha función que transforme o intervalo  $[0, 1]$  en toda a recta real. Así, trataríase de expresar

$$g(\pi(x, \beta)) = x'\beta$$

onde á función  $g$  se lle coñece como función de enlace ou función *link*,  $x$  é o vector que contén os valores das variables explicativas e  $\beta$  é o vector dos coeficientes.

Nesta situación onde a variable resposta é dicotómica, é común considerar como función *link* a función loxística ou función *logit*:

$$g(p) = \log\left(\frac{p}{1-p}\right) \quad \forall p \in [0, 1]$$

Como o argumento  $p$  será substituído pola probabilidade de éxito, a función loxística consiste en efectuar un logaritmo neperiano ó cociente entre a probabilidade de éxito,  $p$ , e a probabilidade de fracaso  $(1 - p)$ . Este cociente é coñecido como *odds* (disparidade ou vantaxe):

$$Odds(Y) = \frac{\mathbb{P}(Y = 1)}{\mathbb{P}(Y = 0)}$$

Agora que a *odds* se move non intervalo  $[0, +\infty]$ , só falta aplicar un logaritmo para transformala nunha cantidade situada en toda a recta real, susceptible de ser explicada mediante un modelo linear. En definitiva, o modelo consistirá en expresar o logaritmo da *odds* da variable resposta como función linear da variable explicativa:

$$\log\left(\frac{\pi(x, \beta)}{1 - \pi(x, \beta)}\right)$$

Para considerar este modelo como representación da probabilidade de éxito, é preciso inverter a función *logit*, quedando así o modelo do seguinte xeito:

$$\pi(x, \beta) = g^{-1}(x'\beta) = \frac{e^{x'\beta}}{1 + e^{x'\beta}}$$

Debido a que a parte linear do modelo se atopa nos expoñentes ós que se eleva o número  $e$ , neste modelo unha variable será pouco significativa se está en torno ó 1.

Para o caso de variables explicativas categóricas, que permiten separar os individuos en dúas ou máis poboacións diferenciadas, defínese *odds ratio* (*OR*) ou cociente de disparidades entre dúas poboacións respecto dunha variable dicotómica  $Y$  como o cociente das *odds* nunha e outra poboación, isto é:

$$OR = \frac{Odds(Y/Pob.2)}{Odds(Y/Pob.1)} = \frac{\mathbb{P}(Y = 1/Pob.2)/\mathbb{P}(Y = 0/Pob.2)}{\mathbb{P}(Y = 1/Pob.1)/\mathbb{P}(Y = 0/Pob.1)} \quad (3.1.1)$$

onde *Pob.1* e *Pob.2* son as dúas poboacións.

### 3.2. Curvas *ROC*

No presente escrito preténdese avaliar e comparar a capacidade diagnóstica de dous modelos diferentes. A curva *ROC* (*Receiver Operating Characteristic*) é unha ferramenta de uso moi estendido para cuantificar o valor diagnóstico dun test dentro do rango de todos os puntos de corte posibles para clasificar os pacientes en positivos ou negativos. Para poder explicar ben a construción de dita curva, é preciso empezar dando unhas nocións básicas.

### 3.2.1. Sensibilidade e especificidade

No caso de que a variable resposta sexa binaria, a precisión dun test resúmese a través das taxas de clasificación correcta, coñecidas como sensibilidade e especificidade.

- **Sensibilidade.** A sensibilidade é a fracción de verdadeiros positivos ( $VP$ ) entre o número total de pacientes que presentan a enfermidade. Un verdadeiro positivo é todo paciente que presenta a enfermidade e que ten un resultado positivo na proba clínica. É dicir, a sensibilidade é a capacidade de clasificar correctamente a un individuo enfermo.

$$\text{Sensibilidade} = \frac{VP}{VP + FN}$$

sendo  $VP$  o número de verdadeiros positivos e  $FN$  o número de falsos negativos (individuos que realmente presentan a enfermidade pero que dan negativo na proba clínica). A sensibilidade é especialmente importante cando a enfermidade non debe pasar desapercibida ou cando o prognóstico mellora moito cun tratamento precoz, como é o caso dalgúns cancros.

- **Especificidade.** A especificidade é a fracción de verdadeiros negativos ( $VN$ ) entre o número total de pacientes que non presentan a enfermidade. Noutras palabras, é a probabilidade de clasificar correctamente a un individuo san.

$$\text{Especificidade} = \frac{VN}{VN + FP}$$

sendo  $VN$  o número de verdadeiros negativos (pacientes que non presentan a enfermidade e obteñen resultados negativos no test) e  $FP$  o número de falsos positivos (pacientes que non presentan a enfermidade e, sen embargo, obteñen un resultado positivo na proba clínica). A especificidade é importante sobre todo cando a enfermidade cambia a calidade de vida dos pacientes ou as consecuencias do tratamento poden supor un risco para o paciente, como unha amputación ou tratamento con radioterapia.

No Cadro 3.2.1 pódese ver a relación entre o resultado da proba diagnóstica e a presenza ou ausencia dunha enfermidade.

Para clasificar un paciente como negativo ou positivo, é preciso establecer un punto de corte ( $c$ ) para a probabilidade predictiva, de xeito que se a predición é máis alta ca  $c$  o paciente será clasificado como positivo e, no caso contrario, como negativo. A representación gráfica dos resultados segundo o punto de corte é a curva *ROC*.

Na Figura 3.2.1 tense a representación gráfica da función de densidade da probabilidade obtida cun modelo diagnóstica para os individuos que non padecen dita enfermidade (sans) e os que si a padecen (enfermos). A modo ilustrativo, estableceuse un punto de corte  $c$  arbitrario. Pódese ver que existe unha zona de posibles resultados da proba para a que a distribucións de suxeitos sans e enfermos se solapan. Ademais, ó mover o punto de corte cara á esquerda, auméntase a probabilidade de detectar pacientes enfermos, aumentando tanto os verdadeiros positivos como os falsos positivos. Pola contra, se se despraza o punto de corte cara á dereita, diminúen os falsos positivos, pero tamén os verdadeiros, aumentando así os falsos negativos. Noutras palabras, ó aumentar a especificidade diminúe a sensibilidade e viceversa.

		Resultados da proba	
		Positivo	Negativo
Realidade	Enfermidade	<b>VP</b>	<b>FN</b>
	Non enfermidade	<b>FP</b>	<b>VN</b>

Cadro 3.2.1: Relación entre o resultado da proba diagnóstica e a ausencia ou presenza dunha enfermidade. Destácanse en diferentes cores os individuos ben clasificados (en negro) dos mal clasificados (en gris). A cantidade de suxeitos realmente enfermos é a suma dos individuos colocados nas celas VP e FN, mentres que os realmente sans obtéñense sumando os individuos das celas VN e FP.

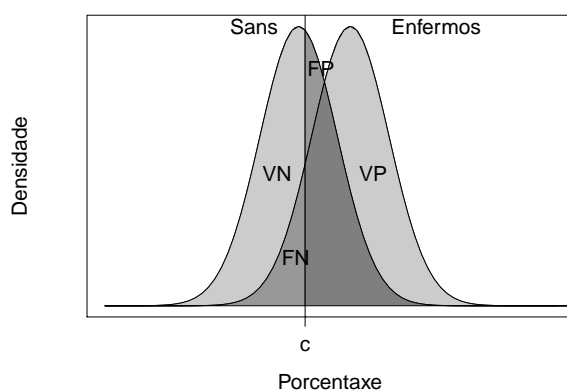


Figura 3.2.1: Representación gráfica das funcións de densidade da probabilidade obtida cun modelo de diagnóstico en función de se os pacientes sofren (enfermos) ou non (sans) dita enfermidade. O punto  $c$  é un punto de corte aleatorio.

### 3.2.2. A curva *ROC*

As curvas *ROC* presentan a *sensibilidade* dunha proba diagnóstica que produce resultados continuos en función dos falsos positivos (complementario da *especificidade*) para distintos puntos de corte. O vértice (0,1), é dicir, o vértice superior esquerdo, representa a clasificación perfecta. Pola contra, a recta con ángulo de 45 graos representa un modelo que non aporta información sobre a clasificación.

Na Figura 3.2.2 tense a representación gráfica dunha curva *ROC*, indicándose tamén o punto que representa o test ideal e a recta que indica que o modelo non aporta ningunha información.

O cálculo dunha curva *ROC* é moi sinxelo e depende do tipo de datos utilizados. É preciso comezar por calcular os puntos de *sensibilidade* e  $1 - \text{especificidade}$  para todos os posibles puntos de corte  $c$ . Se os datos están nunha escala continua, estes puntos sería os valores

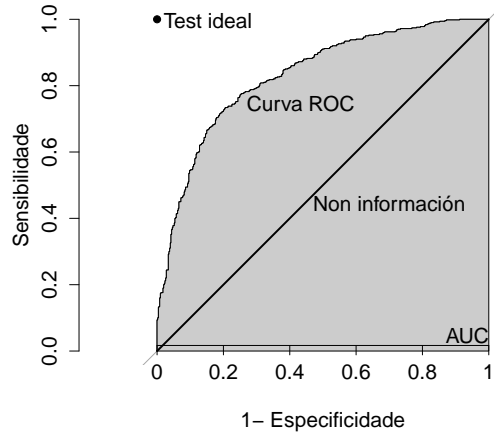


Figura 3.2.2: Representación gráfica dunha curva *ROC*.

observados. Se os resultados foron agrupados en intervalos, estes puntos de corte sería os límites deses intervalos. Por último, se os datos son variables discretas (coma no caso a tratar), os puntos de corte son todos os valores posibles da escala. Os puntos (0,0) e (1,1) sempre pertencen á curva *ROC*.

### Área baixo a curva

A área baixo a curva (*AUC*) pode ser interpretada como a probabilidade de que un paciente coa enfermidade teña unha maior probabilidade segundo o modelo que un paciente aleatorio sen a enfermidade. Este marcador é coñecido como un representante da medida de concordancia entre o marcador e o indicador da enfermidade. A *AUC* defínese como:

$$AUC = \int_0^1 ROC(p)dp \quad (3.2.1)$$

onde  $p \in [0, 1]$  denota a fracción de falsos positivos.

Canto maior sexa a área baixo a curva, mellor discrimina o modelo utilizado. Se este marcador valesse 1, entón a proba sería perfecta e clasificaría á totalidade dos enfermos como enfermos e á totalidade dos sans como tales. Pola contra, se a *AUC* valesse 0.5 (o mínimo valor que pode tomar), só estaría discriminando o 50%, é dicir, habería a mesma probabilidade de clasificar mal a un individuo que de clasificalo ben. Na Figura 3.2.2, a área baixo a curva correspóndese coa parte sombreada.

### 3.3. Avaliación dos modelos de prognóstico

Unha vez estimado un modelo prognóstico, é fundamental realizar unha avaliación do mesmo para ver se é adecuado, cuantificar can boas son as predicións a partir do modelo e para analizar tamén o rendemento dos marcadores. Tradicionalmente, existen múltiples medidas

de avaliación da *performance* do modelo, que se poden agrupar en tres grandes categorías: medidas de axuste xeral, medidas de calibración e medidas de discriminación. Ademais, tamén hai medidas relacionadas coa reclasificación e medidas relacionadas coa utilidade clínica. Para unha revisión en detalle deste tema, pódese consultar Steyerberg (2009) e Steyerberg et al. (2010).

### 3.3.1. Medidas de axuste xeral

Estas medidas céntrase en calcular a distancia entre os valores preditos e os valores reais da variable resposta. No caso de variables continuas, esta distancia é tan só a diferenza entre elas,  $Y - \hat{Y}$ . Para o caso de variables binarias, como é o que se está a tratar,  $\hat{Y}$  é a probabilidade predita  $p$ , mentres que para datos de supervivencia é o tempo predito ata o un evento. É evidente, que para os dous últimos casos, a distancia entre o predito e o observado non se pode calcular da forma indicada, polo que haberá que buscar alternativas. Algunhas das medidas de axuste xeral son as que se amosan a continuación.

#### Varianza explicada: $R^2$ de Nagelkerke

O coeficiente de varianza explicada,  $R^2$  é unha medida global para cuantificar a cantidade de información explicada polo modelo. Esta medida é a máis común para variables resposta continuas, pero non se pode aplicar para modelos con outro tipo de resposta. Para modelos lineares xeralizados, o que se usa é o  $R^2$  de Nagelkerke, que se calcula como segue:

$$R_N^2 = \frac{(1 - \exp(-LR/n))}{(1 - \exp(-2LL_0/n))} \quad (3.3.1)$$

sendo  $n$  o número total de individuos,  $LL_0$  o logaritmo da verosimilitude para o modelo nulo e  $LR$  é a razón de verosimilitude. Tomando  $LL_1$  como o logaritmo da verosimilitude para o modelo que se está a estimar,  $LR = -2(LL_0 - LL_1)$ . Para variables binarias, o logaritmo da verosimilitude para un paciente que de verdade ten a enfermidade codificada mediante a variable resposta é  $\log(p)$ , mentres que, noutro caso, o logaritmo da verosimilitude coincide con  $\log(1 - p)$ , polo que o logaritmo da verosimilitude do modelo será:

$$LL = \sum_{i=1}^n ((Y_i - 1)(\log(1 - p_i)) + Y_i \log(p_i))$$

O  $R^2$  de Nagelkerke toma valores entre 0 e 1, sendo mellor o modelo canto máis grande sexa este valor. Nun modelo perfecto,  $LL_1 = 0$ , polo que  $LR = -2LL_0$ , polo que  $R_N^2 = 1$ .

#### Puntuación de Brier

A puntuación de Brier é unha das medidas globais de avaliación dun modelo máis usadas. Esta medida baséase nas diferenzas cadráticas entre a variable resposta e as predicións  $p$  calculadas.

Este estatístico ben definido da seguinte forma:

$$Brier = \sum_{i=1}^n (Y_i - p_i)^2 = \sum_{i=1}^n (Y_i(1 - p_i)^2 + (1 - Y_i)p_i^2) \quad (3.3.2)$$

Desta forma, cando mellor discrimine o modelo, máis preto estará  $p_i$  de  $Y_i$ , o que se traduce nunha puntuación de Brier máis baixa. Para cada paciente, a puntuación sitúase entre 0 e 1. O mínimo valor que pode tomar a puntuación de Brier é 0, mentres que o máximo depende da incidencia da enfermidade. Se  $inc$  é a incidencia da variable resposta, a máxima puntuación de Brier vén dada por:

$$Brier_{max} = inc(1 - inc)^2 + (1 - inc)inc^2$$

No caso do exemplo que se está tendo en consideración para este traballo, a incidencia é de  $inc = 0.0925$  polo que o valor máximo do índice de Brier (é dicir, un modelo sen capacidade discriminativa) é  $Brier_{max} = 0.084$ . Debido ó reducido tamaño do rango, unha opción é escalar o índice en función do seu valor máximo. Deste xeito, o seu rango estaría entre 0 e 1, moito máis fácil de interpretar.

$$Brier_{escalado} = 1 - \frac{Brier}{Brier_{max}} \quad (3.3.3)$$

### 3.3.2. Medidas de capacidade de discriminación

A discriminación é a capacidade do modelo de asignar o resultado correcto a un par de suxeitos seleccionados aleatoriamente, é dicir, é a habilidade do modelo para distinguir entre os pacientes que teñen a enfermidade en cuestión e os que non.

A medida de discriminación máis usada é a área baixo a curva *ROC* (*AUC*), xa explicado neste mesmo Capítulo 3, pero existen outras medidas que se poden considerar.

#### Diagrama de caixas e pendente de discriminación

A pendente de discriminación é unha medida da separación entre suxeitos que padecen a enfermidade codificada pola variable resposta e os que non. Calcúlase simplemente facendo a diferenza absoluta entre a media das predicións dos individuos dos dous grupos, isto é

$$Pendente = |\mathbb{E}(p_i/Y_i = 1) - \mathbb{E}(p_j/Y_j = 0)| \quad (3.3.4)$$

Unha forma intuitiva de visualizar a separación entre os dous grupos é mediante un diagrama de caixas. Canto mellor discrimine o modelo, máis separación haberá entre as dúas caixas.

### 3.3.3. Avaliación da calibración

A calibración refírese á concordancia entre o observado e o predito e é unha propiedade moi importante dos modelos. Malia que hai varias formas de medir a calibración, a máis importante e, por tanto, a que se vai a tratar neste escrito é o gráfico de calibración.

### Gráfico de calibración

O gráfico de calibración non é máis que a representación dos valores reais da variable resposta fronte ós valores preditos. Predicións perfectas deberíanse situar na recta que forma 45°. Para a regresión linear, este gráfico é un simple gráfico de sedimentación. En cambio, para o caso de variables resposta binarias, as probabilidades non se observan directamente. Neste caso, úsanse técnicas de suavizado para estimar as probabilidades observadas da variable resposta e relación coas probabilidades preditas, combinando para cada individuo os valores reais dos suxeitos con probabilidade predita similar mediante o algoritmo *loess*.

A regresión *loess* (inicialmente *lowess*) é un modelo de regresión que non asume un modelo global fixo, senón de forma local.

Na Figura 3.3.1 pódese ver un exemplo de gráfico de calibración, no que ademais se engadiu o erro cadrático medio.

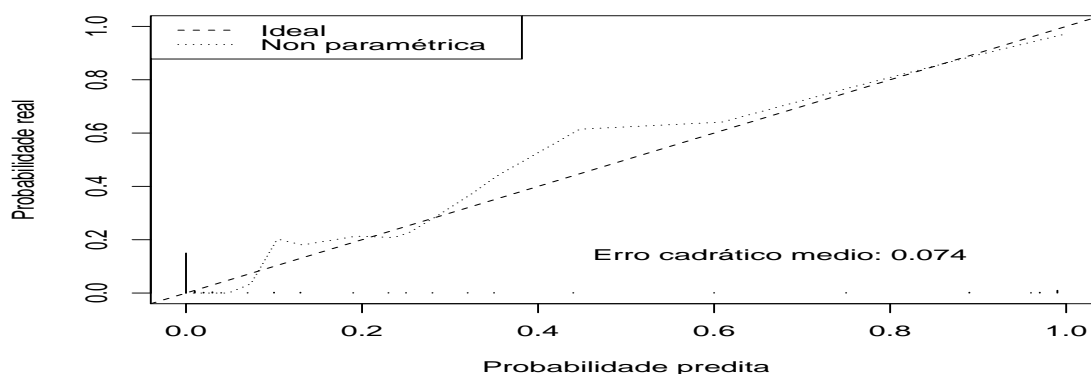


Figura 3.3.1: Exemplo dun gráfico de calibración.

#### 3.3.4. Medidas relacionadas coa reclasificación

Para apreciar as diferenzas entre un modelo e o mesmo pero engadíndolle un marcador máis (como é o primeiro caso que se analiza ó longo deste traballo) existen varias medidas comparativas. A máis usada é a *AUC*, xa tratada anteriormente, pero esta ten o inconveniente de que variacións nalgúns casos son significativas e noutros non, é dicir, non sempre se pode estar seguro se a diferenza de área baixo a curva entre dous modelos é o suficientemente significativa sen realizar algunha proba a maiores, coma un test de hipóteses. Debido a isto, aquí tratarase outro método para tentar ver como cambia a clasificación dos modelos segundo os diferentes modelos, coñecido como mellora neta da reclasificación (*NRI*), que se pode consultar en Pencina et al. (2011).

Aplicar este método non só serve para comparar dous modelos para ver cal é mellor, senón que tamén serve para avaliar un modelo tendo en conta que se pode ampliar. Doutro xeito, para ver se o modelo do que se dispón clasifica o suficientemente ben sen necesidade de engadir



novas variables.

### Mellora neta de reclasificación ou *NRI*

A idea inicial deste método é comparar a clasificación dos individuos segundo cada un dos dous modelos, vendo así cantos individuos cambian de categoría. Pero non só se debería mirar a reclasificación, senón que se debería ter en conta se esta se fixo de forma correcta, é dicir, se un paciente mal clasificado mediante o primeiro modelo está na categoría correspondente mediante o segundo modelo.

O *NRI* ou mellora neta de reclasificación é unha forma sinxela e intuitiva de cuantificar a mellora ofrecida pola nova variable. Noutras palabras, permite cuantificar a cantidade de individuos que están ben reclasificados segundo o modelo ampliado. Debido á súa simpleza, este método foi gañando popularidade entre os investigadores nos últimos anos.

Defínese un movemento ascendente (*up*) como o cambio cara a unha categoría superior e un movemento descendente (*down*) como o cambio a unha categoría inferior. Os movementos ascendentes para os suxeitos que sofren o evento (*event*) implica unha mellora na reclasificación, mentres que un movemento descendente implica un empeoramento. Para os pacientes que non sofren o evento (*non - event*), a interpretación é xusto a oposta. Neste caso, o evento é que o derrame pleural sexa maligno.

O *RNI* defínese como se pode ver en (3.3.5).

$$NRI = \mathbb{P}(up/event) - \mathbb{P}(down/event) + \mathbb{P}(down/non - event) - \mathbb{P}(up/non - event) \quad (3.3.5)$$

*NRI* preséntase así como a suma de dúas compoñentes, unha para os casos de eventos e outra para os non eventos con ponderación proporcional á incidencia do evento.

Aplicando a regra de Bayes, pódese reescribir (3.3.5) dunha forma equivalente:

$$NRI = \frac{\mathbb{P}(event/up)\mathbb{P}(up) - \mathbb{P}(event/down)\mathbb{P}(down)}{\mathbb{P}(event)} \quad (3.3.6)$$

$$+ \frac{\mathbb{P}(non - event/down)\mathbb{P}(down) - \mathbb{P}(non - event/up)\mathbb{P}(up)}{\mathbb{P}(non - event)}$$

Facendo o cambio  $\mathbb{P}(non - event) = 1 - \mathbb{P}(event)$  e  $\mathbb{P}(non - event/*) = 1 - \mathbb{P}(event/*)$ , sendo \* *up* ou *down*, a expresión (3.3.6) queda como segue.

$$NRI = \frac{\mathbb{P}(event/up)\mathbb{P}(up) - \mathbb{P}(event/down)\mathbb{P}(down)}{\mathbb{P}(event)} \quad (3.3.7)$$

$$+ \frac{1 - \mathbb{P}(event/down)\mathbb{P}(down) - (1 - \mathbb{P}(event/up)\mathbb{P}(up))}{1 - \mathbb{P}(event)}$$

A mellora de reclasificación neta pódese interpretar como unha medida do incremento da taxa de evento entre aqueles que foron reclasificados de forma ascendente e da diminución na taxa de evento entre os que foron reclasificados de forma descendente.

Sexa  $n$  o número total de pacientes e sexan  $n_d$  os individuos reclasificados descendentemente e  $n_u$  os suxeitos reclasificados de forma ascendente. Con esta notación, a expresión (3.3.7) queda do seguinte xeito:

$$NRI = \frac{\mathbb{P}(event/up)n_u - \mathbb{P}(event/down)n_d}{n\mathbb{P}(event)} + \frac{1 - \mathbb{P}(event/down)n_d - (1 - \mathbb{P}(event/up)n_u)}{n(1 - \mathbb{P}(event))}$$

Esta fórmula, a diferenza das anteriores, non depende da existencia de categorías de risco xa que as probabilidades de evento entre os individuos cunha reclasificación similar obteranse xuntando todos os individuos co mesmo tipo de reclasificación, é dicir, tense en conta calquera diminución ou aumento na probabilidade predita, independentemente de se é suficiente para un cambio de categoría ou non no caso de habelas. Para mostras grandes, as dúas aproximacións vistas son similares, mentres que para mostras pequenas é preferible esta última. No caso tratado aquí a mostra é suficientemente grande ( $n = 281$ ), polo que se opta por utilizar (3.3.7).

### 3.3.5. Medidas relacionadas coa utilidade clínica

Ademais do visto anteriormente, unha cuestión interesante é a utilidade clínica dun modelo, é dicir, os beneficios que aporta dito modelo na práctica clínica para determinar un diagnóstico ou determinado tratamento. Neste contexto, é preciso establecer un punto de corte de xeito que os individuos con probabilidade superior son clasificados como positivos, mentres que se clasifican como negativos se ocorre o contrario. Aquí, defínese utilidade clínica como a habilidade dun modelo para facer dita clasificación dun mellor xeito que cunha norma establecida sen ter en conta o modelo.

Unha cuestión vital para clasificar os suxeitos en función da probabilidade predita como enfermos ou non enfermos (que presentan o evento ou non o presentan) é o punto de corte. Tradicionalmente, o punto de corte era 0.5, é dicir, que teñen a mesma importancia os falsos negativos coma os falsos positivos, pero isto non é o máis usual en medicina. Pode ocorrer que teña máis peso o feito de clasificar erroneamente a un paciente que non presente o evento ca non diagnosticar a un paciente que si que o presente. Esta situación ocorre sobre todo cando o tratamento é moi agresivo, polo que convén aplicarllo só a persoas que realmente o precisen (por exemplo, unha amputación). Por outra banda, tamén é posible que perder un paciente que presente o evento sexa peor ca clasificar coma positivo alguén que non debería selo. Isto é o usual cando se trata de enfermidades moi agresivas, como é o caso do cancro. Este último suposto é o que sucede cos pacientes de derrame pleural, no que é moito máis importante evitar que un paciente cuxa causa do derrame sexa maligna sexa clasificado como negativo ca clasificar como positivo un individuo con causa non cancerosa.

Na práctica, determinar o punto de corte óptimo é unha cuestión moi complicada, xa que pode depender da poboación usada para derivar o modelo, así coma do propio paciente, xa que os riscos e os beneficios dende o seu punto de vista cambian en función dun suxeito a outro. Como este punto de corte no é fixo, pode ser que o modelo utilizado non presente un gran avance á hora de clasificar pacientes. Para determinar esta cuestión existen varios métodos, destacando entre eles a curva de decisión (para máis información, pódese consultar Vickers e Elkin (2006)).

### Curvas de decisión

A curva de decisión é unha ferramenta moi útil á hora de escoller un modelo de diagnóstico clínico en función do punto de corte escollido. Consiste na representación gráfica, para diferentes puntos de corte situados entre 0 e 1, do coñecido como beneficio neto.

Para avaliar un modelo, emprégase a relación existente entre o punto de corte da probabilidade de enfermidade e valor relativo de falsos negativos ( $FN$ ) e falsos positivos ( $FP$ ). Entón, para un determinado punto de corte  $c$ , o beneficio neto é:

$$BN = \frac{VP}{n} - \frac{FP}{n} \left( \frac{c}{1-c} \right) \quad (3.3.8)$$

Segundo a fórmula anterior, o beneficio neto non é máis que a diferenza entre a taxa de verdadeiros positivos e a taxa de falsos negativos ponderada polo dano relativo dos falsos negativos e dos falsos positivos. Canto mellor sexa un modelo, maior beneficio neto xa que hai un maior número de verdadeiros positivos e unha menor cantidade de falsos positivos. Teoricamente, o rango do beneficio neto vai desde  $-\infty$  ata a incidencia da enfermidade.

Para determinar se o modelo ten un beneficio neto alto, é preciso comparalo con outros dous modelos: o modelo nulo, que consiste en supoñer que ningún paciente presenta a enfermidade (é dicir, todos negativos) e que leva a un beneficio neto de 0 ó non haber ningún tipo de positivos, e o modelo "tratar a todos", que consiste en supoñer que todos os individuos presentan dita enfermidade (isto é, todos positivos).

A interpretación do beneficio neto faise en unidades de verdadeiros positivos, isto é, cantos máis pacientes que presentan o evento están correctamente clasificados sen incrementar taxa de pacientes non enfermos clasificados como positivos.

A redución no número de falsos positivos por cada 100 individuos calcúlase como segue:

$$\frac{BN_{modelo} - BN_{tratar\ a\ todos}}{c/(1-c)} 100 \quad (3.3.9)$$

Na Figura 3.3.2 pódese ver un exemplo de curva de decisión.

No exemplo considerado na Figura 3.3.2 só se consideraron probabilidades menores a 0.5. A liña gris é o beneficio neto se se consideran todos os pacientes negativos, mentres que a liña continua toda é o modelo "tratar a todos", é dicir, considerando que todos son positivos. O punto onde se cortan estas dúas curvas é a incidencia da enfermidade.

## 3.4. Presentación

Como se trata de modelos de predición clínica que van ser usados polos médicos, é preciso presentar os seus resultados de forma que sexan sinxelos de interpretar a aplicar a novos casos. Aínda que existen varios métodos, aquí vaise tratar o nomograma.

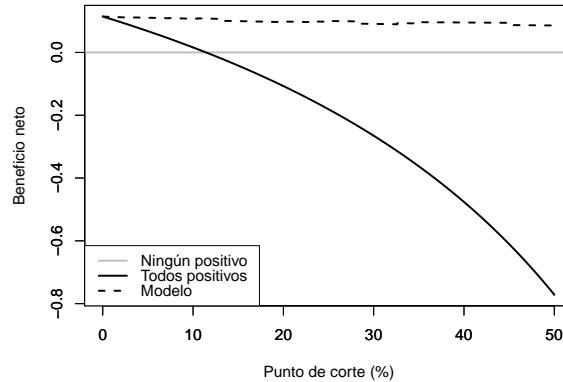


Figura 3.3.2: Exemplo dunha curva de decisión, con  $c \in [0, 0.5]$ .

### Nomogramas

Os nomogramas son representacións gráficas dun modelo de predición. Na Figura 3.4.1 tense un exemplo deste tipo de gráficos.

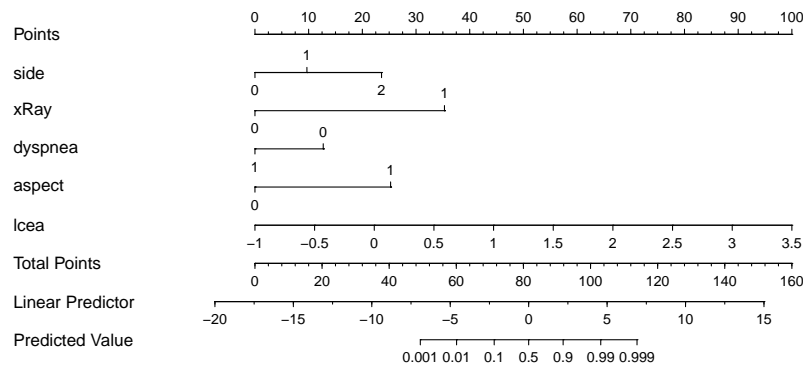


Figura 3.4.1: Exemplo de nomograma.

Na Figura 3.4.1, a liña *Points* é a liña de referencia das puntuacións, normalmente de 0 a 10 ou de 0 a 100. As liñas seguintes representan cada unha das variables do modelo con cada un dos seus posibles valores, de forma que se poden saber os puntos correspondentes a cada valor tan só comparándoo coa liña de referencia *Points*. Unha vez sumados todos os puntos dun paciente, é posible situar a súa puntuación total na liña *Total Points*. Na liña inmediatamente inferior tense o *Linear Predictor* ou predictor linear, véndose así a equivalencia entre os puntos totais de cada individuo e o seu valor do predictor linear do modelo. Finalmente, na liña inferior, téñense os valores preditos. Para ver a correspondencia entre as predicións e o predictor linear, non hai máis que trazar unha liña vertical. Por exemplo, un valor predito de 0.5 correspóndese cun predictor linear de 0. É habitual acompañar os nomogramas dunha táboa na que se expresen as

puntuacións de cada variable de forma exacta, ademais de ver a cantos puntos corresponden as probabilidades habituais, como son 0.01 ou 0.1. A continuación pódense ver os cadros resumo do nomograma da Figura 3.4.1.

```
## Points per unit of linear predictor: 6.773632
## Linear predictor units per point   : 0.1476313
##
##
## side Points
## 0      0
## 1     15
## 2     35
##
##
## xRay Points
## 0      0
## 1     56
##
##
## dyspnea Points
## 0      21
## 1       0
##
##
## aspect Points
## 0      0
## 1     38
##
##
## lcea Points
## -1.0   0
## -0.5  11
##  0.0  22
##  0.5  33
##  1.0  44
##  1.5  56
##  2.0  67
##  2.5  78
##  3.0  89
##  3.5 100
##
##
## Total Points Predicted Value
##           65           0.001
##           80           0.010
##           96           0.100
```

##	111	0.500
##	126	0.900
##	142	0.990
##	158	0.999

As dúas primeiras liñas indican a equivalencia entre os puntos e o valor do predictor linear. Así, por exemplo, se un paciente ten derrame no pulmón esquerdo ( $side = 2$ ), sen presenza de elementos suxestivos de malignidade nunha proba imaxe ( $xRay = 0$ ), sen dispnea ( $dyspnea = 0$ ), con líquido de aspecto suxo ( $aspect = 1$ ) e con  $lcea = 0.5$  terá 95 puntos, o que se corresponde segundo a última táboa con case 0.1 de probabilidade de que o derrame pleural sexa maligno.

### 3.5. Corrección do sobreaxuste e optimismo

Para construír un modelo empréganse datos empíricos procedentes dunha mostra de individuos, á súa vez procedente dunha poboación. O problema é que as predicións non se poden xeralizar a novos individuos fóra da mostra, xa que se corre o risco de que estas dependan de aspectos particulares da mesma que non están presentes na poboación total. A este fenómeno chámase sobreaxuste. En modelos de predición, o sobreaxuste defínese como o axuste dun modelo estatístico con demasiados graos de liberdade e conduce a unha impresión demasiado optimista da *performance* do modelo. O optimismo defínese como a diferenza entre o verdadeiro rendemento do modelo e o rendemento aparente.

Para finalizar, levaranse a cabo varias accións para intentar calcular o optimismo e tentar corrixir o sobreaxuste, entre as que se inclúe a corrección dos coeficientes dos modelos segundo diferentes métodos. Esta sección estará centrada na corrección da *AUC*.

#### Cálculo da *performance* corrixida de optimismo

Unha aplicación moi importante da metodoloxía *bootstrap* é a cuantificación do optimismo dun modelo de predición. Para cada mostra *bootstrap*, axústase un modelo que logo se avaliará cos datos da mostra orixinal, calculándose o optimismo como a diferenza entre a *performance* coas dúas mostras. A estimación da *performance* corrixida de optimismo non é máis que a medida de validación inicial sen o optimismo calculado.

#### *Shrinkage* uniforme

O encollemento dos coeficientes dun modelo a cero, ou *shrinkage*, é unha forma de mellorar as predicións dun modelo de regresión. Existen varias formas, destacando o *shrinkage* uniforme.

Os coeficientes encollidos so calculados simplemente como  $s\beta$ , onde  $s$  é o factor de encollemento uniforme e  $\beta$  son os coeficientes de regresión estimados. Existe unha fórmula heurística para

calcular o  $s$ :

$$s = \frac{(\chi_{\text{modelo}}^2 - df)}{\chi_{\text{modelo}}^2}$$

sendo  $-\chi_{\text{modelo}}^2$  a razón de verosimilitude do modelo axustado e  $df$  os graos de liberdade dos preditores considerados para o modelo.

### **Lasso**

A metodoloxía *Lasso* (*least absolute shrinkage and selection operator*) estima os coeficientes dos preditores estandarizados mediante a minimización do logaritmo da verosimilitude suxeito a  $\sum |\beta| \leq t$ . Neste caso,  $t$  determina o encollemento no modelo. Os coeficientes finais estimaranse co valor de  $t$  que proporcione un erro cadrático medio menor nun procedemento de validación cruzada xeralizada.

Os modelos cuxos coeficientes foron modificados mediante estas dúas metodoloxías, *shrinkage* uniforme e *Lasso*, pódense avaliar da mesma forma que se explicou anteriormente, aínda que neste escrito se fará só para a área baixo a curva ou *AUC*.

## **3.6. Resultados para os datos de derrame pleural**

O obxectivo deste apartado é avaliar e comparar dous modelos de regresión que permitan discernir, entre os casos de derrame pleural transudado, aqueles que poidan ter una causa maligna, véxase cancro. As variables a considerar son as expostas no Apartado 2.1.1 do presente escrito, considerándose primeiro un modelo só con variables clínicas e radiolóxicas, para logo considerar outro modelo coas mesmas variables, ademais dun marcador tumoral.

O primeiro paso é atopar un modelo loxístico das variables clínicas. Para iso comezouse axustando un modelo con todas as variables clínicas dispoñibles e realizando eliminación *backward* usando como criterio de parada aquel modelo que minimize o Criterio de Información de Akaike (*AIC*), o cal se define do seguinte xeito:

$$AIC = -2 \ln(\text{verosimilitude}) + 2k$$

onde  $k$  é o número de parámetros a estimar no modelo. Este criterio consiste, pois, en elixir aquel modelo que maximice a verosimilitude pero cun número reducido de parámetros.

Deste xeito, chegamos ó modelo que conta só, como variables explicativas, co lado no que se produciu o derrame, os resultados das probas de imaxe, a presenza de dispnea, se o paciente ten ou non dor e o aspecto do líquido. Ó analizar a significación de cada variable, hai dúas variables que non son significativas no modelo (*pain* e *dyspnea*). Ademais, o *AIC* é 72.67. Tendo en conta isto e que o *AIC* do modelo sen a variable *pain* (a de maior *p* – *valor*) é de 72.83, é dicir, un pouco maior pero sen grandes diferenzas, óptase por quitar dita variable do modelo por cuestión de simplicidade. Por tanto, o primeiro modelo a considerar (*mod1*) será o modelo que se corresponde co Cadro 3.6.1.

Do Cadro 3.6.1 extráese que non hai diferenzas significativas á hora de presentar derrame pleural maligno entre os pacientes con derrame nos dous pulmóns (*side* = 0, a categoría que

Variable	Coefficiente	Erro estándar	$p$ – valor	Odds ratio	OR inf. 95 %	OR sup. 95 %
Intercepto	-6.025	1.413	<0.001	-	-	-
$side = 1$	0.207	0.833	0.803	1.230	0.240	6.297
$side = 2$	2.455	0.959	0.011	11.643	1.777	76.307
$xRay = 1$	6.120	1.220	<0.001	454.700	41.644	4964.800
$dyspnea = 1$	-1.666	0.774	0.031	0.189	0.041	0.862
$aspect = 1$	3.731	1.113	<0.001	41.722	4.710	369.580

Cadro 3.6.1: Variables incluídas no modelo *mod1*, xunto coa súa estimación, o seu erro estándar, o seu  $p$  – valor, a súa *Odds ratio* e intervalo de confianza ó 95 % para a mesma.

se toma de referencia) e ter o derrame só no pulmón dereito. En cambio, se o derrame se produce só no pulmón esquerdo, as probabilidades de que sexa maligno son 11.643 as que probabilidades no caso de que se produza nos dous. Como cabía esperar, atopar elementos suxestivos de malignidade nas probas de imaxe aumenta moito o risco de que a causa sexa cancerosa (*Odds ratio* de 454.700), así como que o aspecto do fluído sexa menos claro (*Odds ratio* de 41.722). Pola contra, a presenza de dispnea no paciente parece ser un factor protector, xa que os pacientes que sofren deste mal teñen 0.189 veces as probabilidades de que o derrame sexa maligno que teñen as persoas que non padecen dispnea. Na Figura 3.6.1, esquerda, tense a representación gráfica da *Odds ratio* de cada variable (ver (3.1.1)) co seu respectivo intervalo de confianza.

Unha vez considerado este modelo, procédese ó axusto do modelo *mod2*, o cal conta coas mesmas variables e, a maiores, o *lcea*, é dicir o logaritmo do marcador tumoral CEA. Aínda que nun principio se pensou en aplicar suavización mediante *splines* a esta variable, decidiuse axustar o modelo considerándoa linear debido a que o obxectivo é crear unha regra diagnóstica que poida servir ó médicos no seu traballo, polo que lles pode resultar complicado de entender e aplicar se se introducen *splines* e variables suavizadas no modelo. O axuste do modelo *mod2* pódese ver no Cadro 3.6.2.

No Cadro 3.6.2 pódese ver que o efecto das variables comúns ós dous modelos segue na mesma liña, pero variando a súa magnitude. Por exemplo, o feito de atopar elementos suxestivos de malignidade nas probas de imaxe pasa de aumentar o risco de que a causa sexa cancerosa 454.700 veces a aumentalo case 70000 veces. En canto á nova variable, pódese ver que o aumento do *lcea* produce tamén un aumento no risco de que o derrame sexa maligno. Como é unha variable continua, a *Odds ratio*, só se pode calcular comparando dúas cantidades concretas, polo que aquí se comparaba o primeiro cuartil ( $lcea = -0.301$ ) co terceiro cuartil ( $lcea = 0.23040$ ), dando idea de que ter o logaritmo do CEA por riba de 0.23 provoca ter case 10 veces a probabilidade de malignidade ca se dito valor está en torno a -0.3. Na Figura 3.6.1, dereita,



Variable	Coefficiente	Erro estándar	$p$ – valor	Odds ratio	OR inf. 95 %	OR sup. 95 %
Intercepto	-13.172	4.015	0.001	-	-	-
$side = 1$	2.033	1.520	0.181	7.639	0.388	150.220
$side = 2$	5.576	2.225	0.012	264.07	3.374	20667
$xRay = 1$	11.155	3.347	<0.001	69906	99.066	$4.9329 \cdot 10^7$
$dyspnea = 1$	-3.584	1.528	0.019	0.028	0.001	0.555
$aspect = 1$	7.929	2.530	0.002	2775.9	19.491	$3.9534 \cdot 10^5$
$lcea$	4.327	1.350	0.001	9.969	2.444	40.668

Cadro 3.6.2: Variables incluídas no modelo  $mod2$ , xunto coa súa estimación, o seu erro estándar, o seu  $p$  – valor, a súa Odds ratio e intervalo de confianza ó 95 % para a mesma.

pódese ver a representación gráfica das Odds ratios do modelo  $mod2$ , xunto cos seus intervalos de confianza.

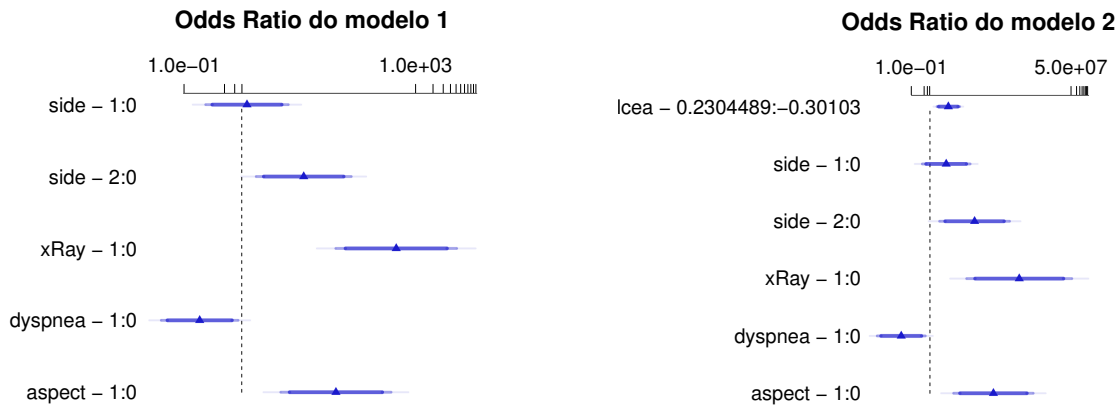


Figura 3.6.1: Representación gráfica das Odds ratios dos dous modelos estudados xunto cos seus intervalos de confianza. Esquerda:  $mod1$ . Dereita:  $mod2$ .

Con afán de ver a capacidade discriminatória de cada modelo, ademais de comparalos entre eles, obtivéronse as curvas ROC correspondentes a cada un e calculáronse as correspondentes áreas baixo a curva, ou AUC (ver (3.2.1)). Na Figura 3.6.2 pódense apreciar nos extremos a curva ROC de cada modelo coas bandas de confianza ó 95 % e coa AUC correspondente a cada unha, tamén con intervalo de confianza ó 95 %. Na imaxe central aparecen as dúas curvas ROC, o que permite apreciar ben a diferenza entre elas. Cabe destacar que calquera

dos dous modelos ten unha gran capacidade discriminatória, xa que as súas áreas baixo a curva superan o 0.97, polo que clasifican correctamente a máis do 97% dos individuos. Outra cousa a destacar é que o modelo 2 ten maior  $AUC$  (como era de esperar) e o seu intervalo de confianza chega ata 1. Isto quere dicir que para calquera modelo que se poia elixir cun maior  $AUC$  ca o modelo 2, non se atoparán diferenzas significativas, polo que se trata dun modelo practicamente óptimo.

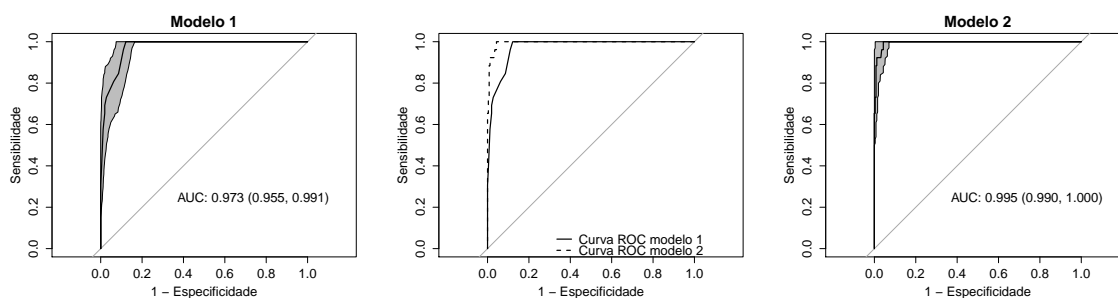


Figura 3.6.2: Curva  $ROC$  correspondentes ós modelos  $mod1$  e  $mod2$  cos seus correspondentes  $AUC$ . Esquerda: modelo 1. Centro: Modelo 1 (liña continua) e modelo 2 (liña discontinua). Dereita: modelo 2.

Ademais, diso, mediante técnicas de remostraxe *wild bootstrap* con 1000 mostras *bootstrap*, achouse un intervalo de confianza ó 95% para a diferenza das  $AUC$ 's dos modelos. Teoricamente, se este intervalo contén o 0, non existen evidencias significativas de diferenzas entre as dúas curvas, mentres que, en caso contrario, rexéitase a hipótese inicial de igualdade, téndose entón que aínda que parece haber pouca diferenza entre as dúas áreas baixo a curva, o modelo 2 discrimina significativamente mellor. No Cadro 3.6.3 recóllense os resultados de dito contraste.

$AUC$	Diferenza das $AUC$ 's	$AUC$ inf. 95%	$AUC$ sup. 95%
Modelo 1: 0.973	0.022	0.006	0.033
Modelo 2: 0.995			

Cadro 3.6.3: Resultados do contraste de hipóteses con hipótese nula a igualdade das  $AUC$ 's dos modelos  $mod1$  e  $mod2$ .

Do Cadro 3.6.3 dedúcese que, aínda que os dous modelos son moi bos á hora de discriminar entre derrame pleural ou non maligno, existen diferenzas significativas entre os dous modelos, sendo mellor o modelo 2.

Chegados a este punto, é vital avaliar os modelos. Para iso, comezase primeiro por medidas de avaliación global dos mesmos. No Cadro 3.6.4 recóllense as medidas mencionadas anteriormente neste Capítulo 3, é dicir, o  $R^2$  de Nagelkerke, a puntuación de Brier e a puntuación de Brier escalada (ver (3.3.1), (3.3.2) e (3.3.3)).

	$R^2$ de Nagelkerke	Puntuación de Brier	Puntuación de Brier escalada
Modelo 1	0.717	0.035	0.583
Modelo 2	0.879	0.015	0.821

Cadro 3.6.4: Medidas de avaliación xeral para os modelos *mod1* e *mod2*.

Como xa se deducira do valor da  $AUC$  para cada un dos modelos (ver Cadro 3.6.3), os resultados do Cadro 3.6.4 indican un mellor axuste do modelo 2 (o que inclúe o logaritmo do marcador tumoral CEA) ca do modelo 1. Tamén cabe destacar que a puntuación de Brier indica un menor axuste ca o  $R_N^2$ .

Unha vez comprobado o axuste global dos modelos, convén avaliar a súa capacidade de discriminación. Na Figura 3.6.3 pódense observar dous diagramas de caixas, o da esquerda correspondente ó modelo 1 e o da dereita, ó modelo 2. Estes diagramas de caixas representan a probabilidade predita polo modelo en función da variable resposta. Na parte superior indícase a súa pendente de discriminación (ver (3.3.4)), sendo considerablemente maior a pendente de discriminación do modelo 2, como cabía esperar. Esta medida, aínda que dá unha idea da capacidade discriminativa do modelo, non é tan ilustrativa como a  $AUC$ , xa que, por exemplo, para o primeiro modelo se dá un valor moito máis baixo, o que podería levar a pensar que o modelo non é moi bon á hora de discriminar, cando se veu anteriormente que ocorre todo o contrario.

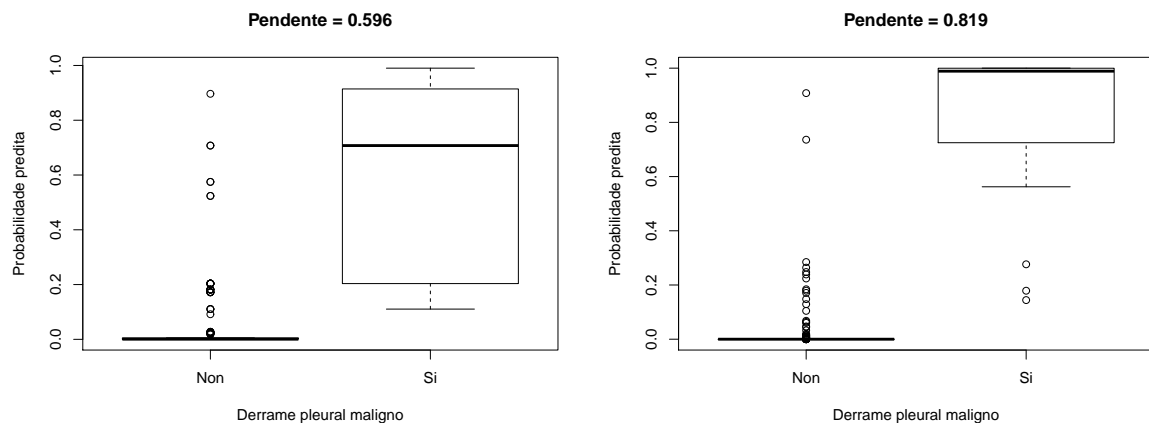


Figura 3.6.3: Diagrama de caixas da probabilidade  $p$  predita polo modelo en función da presenza ou ausencia de derrame pleural maligno, coa pendente de discriminación. Esquerda: modelo 1. Dereita: modelo 2.

Da Figura 3.6.3 extráese o feito de que os dous modelos discriminan bastante ben, algo que xa se vira coa  $AUC$  (ver Cadro 3.6.3).

Para comprobar a calibración dos modelos, na Figura 3.6.4 téñense os gráficos de calibración

dos mesmos.

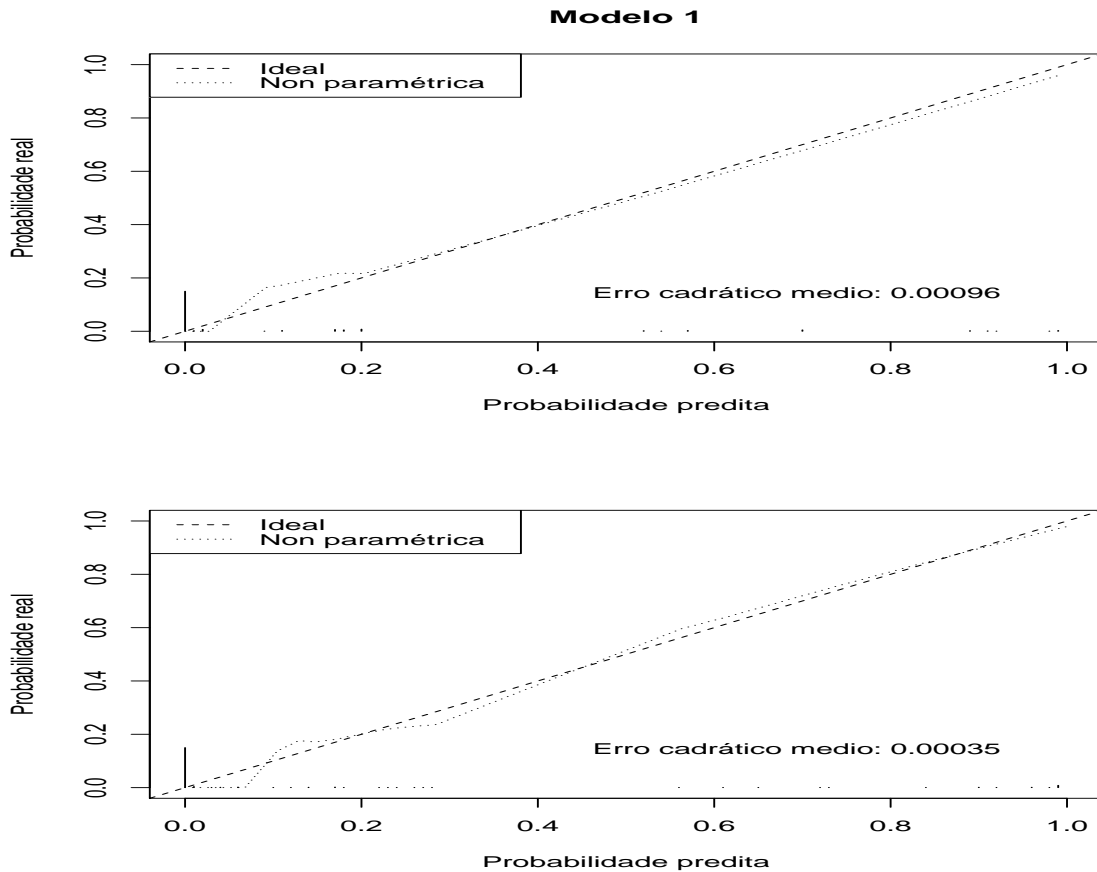


Figura 3.6.4: Gráficos de calibración dos modelos estudados. Arriba: modelo 1. Abaixo: modelo 2.

Tal e como se pode ver na Figura 3.6.4, as predicións de ámbolos dous modelos concordan moi ben coas observacións (ver liña discontinua máis fina). Ademais, para valores pequenos de probabilidade, o modelo 1 leva a unha infraestimación, xa que son menores os valores preditos que a estimación das probabilidades observadas. En cambio, para o modelo 2, ata unha probabilidade de 0.45 aproximadamente, o modelo presenta tramos de infraestimación e de sobreestimación. Aínda así, estas variacións con respecto ó ideal non son de gran magnitude.

Unha vez avaliados os modelos de xeito tradicional, realízase unha avaliación para ver a mellora do modelo 1 ó engadirlle a variable *lcea* (modelo 2) mediante a mellora de reclasificación neta (*RNI*) para dúas categorías de risco (non ter derrame pleural maligno e si ter derrame pleural maligno). Como o modelo loxístico estima probabilidades de ter ou non o evento, é preciso establecer un punto de corte a partir do cal un individuo se clasifica como  $mpe = 1$ . Debido á gravidade do evento, o punto de corte é unha cifra baixa, establecida 0.05 polo pneumólogo investigador. Empregando entón (3.3.7) tense:

$$NRI = 0.063$$

Isto quere dicir que hai un 6.3% dos individuos que estaban mal clasificados co primeiro modelo pero que ó aplicar o modelo 2 xa están na súa categoría correspondente. Aínda que o *NRI* pode parecer pequeno (o cal ten lóxica tendo en conta a boa calibración do modelo), é superior á diferenza de *AUC* vista na Figura 3.6.3 (0.022), dando así unha maior noción de significación.

Comparando agora a utilidade clínica dos dous modelos, na Figura 3.6.5 téñense dous gráficos. O gráfico da esquerda inclúe as curvas de decisión dos dous modelos, ademais do modelo que consiste en considerar todos os pacientes como positivos (liña negra continua) e como negativos (liña gris). Pódese ver neste gráfico que o modelo 2 ten un maior beneficio neto ca o modelo 1. Para ver se esta diferenza de beneficio neto é significativa, mediante técnicas de remostraxe *wild bootstrap* con 1000 mostras *bootstrap*, achouse un intervalo de confianza ó 95% para a diferenza do beneficio neto dos dous modelos. Teoricamente, se este intervalo contén o 0, non existen evidencias significativas de diferencias entre as dúas curvas, mentres que, en caso contrario, rexéitase a hipótese inicial de igualdade. Pódese ver, que ata  $c = 0.05$ , non se atopan diferenzas significativas entre os dous modelos. O beneficio neto foi calculado seguindo (3.3.8) para puntos de corte  $c \in [0, 0.2]$  por recomendación dos clínicos.

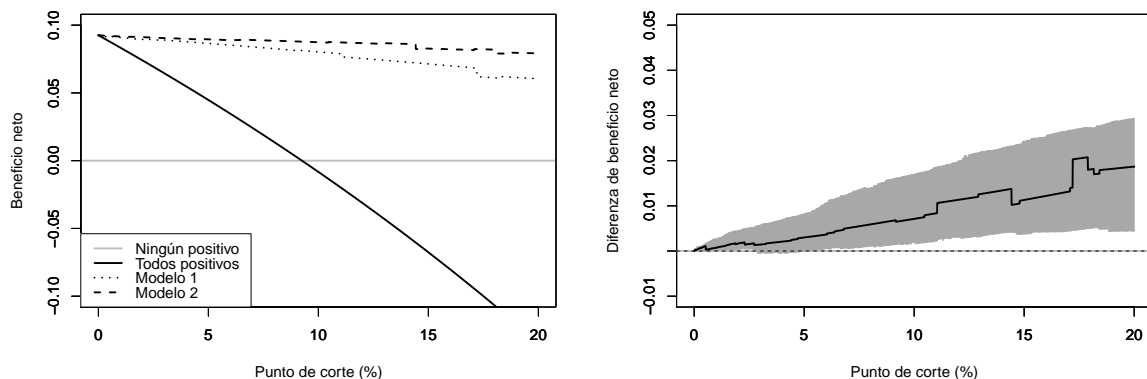


Figura 3.6.5: Esquerda: Curvas de decisión para os modelos *mod1* e *mod2*. Dereita: Gráfico coa diferenza das curvas de decisión dos dous modelos, así coma o seu intervalo de confianza ó 95%.

Os Cadros 3.6.5 e 3.6.6 son dous cadros resumo comparativos dos modelos *mod1*, *mod2* e o modelo “tratar a todos”, nos que se comparan os modelos entre si segundo (3.3.9) para os distintos puntos de corte  $c \in [0, 0.20]$ .

Unha vez validado o modelo, convén presentar os resultados dunha forma práctica e sinxela, pensando en que son os médicos os que van utilizar estes modelos para diagnosticar os pacientes. Para iso imos usar os xa explicados nomogramas. Na Figura 3.6.6 pódese ver o nomograma do modelo 1.

Na Figura 3.6.6 apréciase que a presenza de elementos suxestivos de malignidade nas probas de imaxe ( $xray = 1$ ) xa suma 100 puntos, polo que cabe pensar que se un suxeito nesas circunstancias, independentemente das demais variables, será clasificado como positivo.

<i>c</i> %	Beneficio neto			<i>model1</i>		<i>model2</i>	
	“Tratar a todos”	<i>model1</i>	<i>model2</i>	<i>FP</i> (%)	<i>FN</i> (%)	<i>FP</i> (%)	<i>FN</i> (%)
1	0.083	0.091	0.092	45 (0.634)	0 (0.000)	25 (0.490)	0 (0.000)
2	0.074	0.089	0.091	42 (0.618)	0 (0.000)	19 (0.422)	0 (0.000)
3	0.064	0.089	0.090	32 (0.552)	0 (0.000)	19 (0.422)	0 (0.000)
4	0.055	0.088	0.090	32 (0.552)	0 (0.000)	18 (0.409)	0 (0.000)
5	0.045	0.087	0.090	32 (0.552)	0 (0.000)	16 (0.381)	0 (0.000)
6	0.035	0.085	0.089	32 (0.552)	0 (0.000)	16 (0.381)	0 (0.000)
7	0.024	0.084	0.089	32 (0.552)	0 (0.000)	13 (0.333)	0 (0.000)
8	0.014	0.083	0.089	32 (0.552)	0 (0.000)	13 (0.333)	0 (0.000)
9	0.003	0.081	0.088	32 (0.552)	0 (0.000)	13 (0.333)	0 (0.000)
10	-0.008	0.080	0.087	31 (0.544)	0 (0.000)	13 (0.333)	0 (0.000)
11	-0.020	0.079	0.087	31 (0.544)	0 (0.000)	12 (0.316)	0 (0.000)
12	-0.031	0.075	0.087	28 (0.528)	1 (0.004)	12 (0.316)	0 (0.000)
13	-0.043	0.074	0.087	28 (0.528)	1 (0.004)	11 (0.297)	0 (0.000)
14	-0.055	0.073	0.086	28 (0.528)	1 (0.004)	11 (0.297)	0 (0.000)
15	-0.068	0.071	0.083	28 (0.528)	1 (0.004)	10 (0.286)	1 (0.004)
16	-0.080	0.070	0.082	28 (0.528)	1 (0.004)	10 (0.286)	1 (0.004)
17	-0.093	0.069	0.082	28 (0.528)	1 (0.004)	10 (0.286)	1 (0.004)
18	-0.107	0.061	0.079	22 (0.500)	4 (0.017)	8 (0.250)	2 (0.008)
19	-0.120	0.061	0.080	16 (0.432)	5 (0.020)	7 (0.226)	2 (0.008)
20	-0.134	0.060	0.079	16 (0.432)	5 (0.020)	7 (0.226)	2 (0.008)

Cadro 3.6.5: Cadro comparativo do modelo 1, modelo 2 e modelo “tratar a todos” segundo os seu beneficio neto.

c%	Vantaxe <i>modell</i> vs. “tratar a todos”		Vantaxe <i>model2</i> vs. “tratar a todos”		Vantaxe <i>modell</i> vs. <i>model2</i>	
	Beneficio neto	Redución <i>FP</i>	Beneficio neto	Redución <i>FP</i>	Beneficio neto	Redución <i>FP</i>
1	0.008	79	0.009	89	0.001	10
2	0.015	74	0.017	83	0.002	10
3	0.025	81	0.026	84	0.001	3
4	0.033	79	0.035	84	0.002	5
5	0.042	80	0.045	86	0.003	6
6	0.050	78	0.054	85	0.004	6
7	0.060	80	0.065	86	0.005	7
8	0.069	79	0.075	86	0.006	7
9	0.078	79	0.085	86	0.007	7
10	0.088	79	0.095	86	0.007	6
11	0.099	80	0.107	87	0.008	6
12	0.106	78	0.118	87	0.012	9
13	0.117	78	0.130	87	0.013	9
14	0.128	79	0.141	87	0.013	8
15	0.139	79	0.151	86	0.012	7
16	0.150	79	0.162	85	0.012	6
17	0.162	79	0.175	85	0.013	6
18	0.168	77	0.186	85	0.018	8
19	0.181	77	0.200	85	0.019	8
20	0.194	78	0.213	85	0.019	8

Cadro 3.6.6: Cadro comparativo do modelo 1, modelo 2 e modelo “tratar a todos” segundo os seu beneficio neto. Redución *FP* indica o número de casos de pacientes sen o evento que eran clasificados como positivos segundo o primeiro modelo e como negativos segundo o segundo modelo, por cada 100 pacientes.

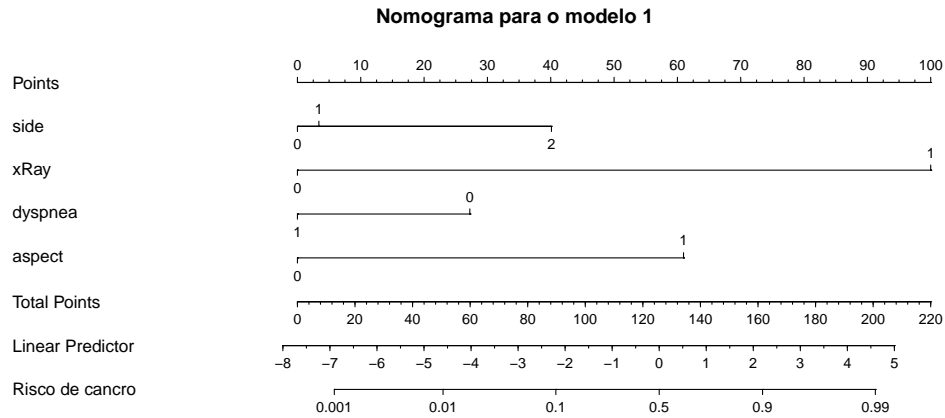


Figura 3.6.6: Nomograma para o modelo 1.

A continuación móstranse as táboas coas puntuacións para cada variable e coa correspondencia entre puntuacións e probabilidades para o modelo 1.

```
## Points per unit of linear predictor: 16.34082
## Linear predictor units per point   : 0.06119644
##
##
## side Points
## 0     0
## 1     3
## 2    40
##
##
## xRay Points
## 0     0
## 1    100
##
##
## dyspnea Points
## 0     27
## 1      0
##
##
## aspect Points
## 0     0
## 1    61
##
##
## Total Points Risco de cancro
```



##	13	0.001
##	51	0.010
##	90	0.100
##	126	0.500
##	162	0.900
##	201	0.990

Así, segundo este modelo, un paciente con derrame pleural no lado dereito ( $side = 1$ ), sen presenza de elementos suxestivos de malignidade nas probas de imaxe ( $xRay = 0$ ), sen dispnea ( $dyspnea = 0$ ) e con líquido pleural sucio ( $aspect = 1$ ) suma 91 puntos, polo que a probabilidade de causa cancerosa é lixeiramente superior a 0.1.

Con respecto ó modelo 2, aquel que inclúe o logaritmo do marcador tumoral CEA, a Figura 3.6.7 mostra o nomograma de dito modelo. Ademais, a continuación móstranse as táboas coas puntuacións das variables do modelo 2, así como a correspondencia entre puntos e probabilidade predita.

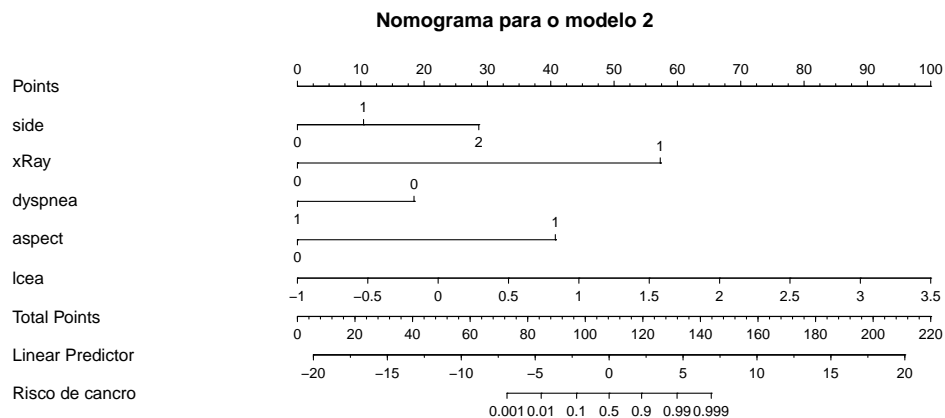


Figura 3.6.7: Nomograma para o modelo 2.

```
## Points per unit of linear predictor: 5.136279
## Linear predictor units per point   : 0.1946935
##
##
## side Points
## 0    0
## 1    10
## 2    29
##
##
## xRay Points
## 0    0
```

```

## 1    57
##
##
##  dyspnea Points
##  0     18
##  1     0
##
##
##  aspect Points
##  0     0
##  1    41
##
##
##  lcea Points
## -1.0  0
## -0.5  11
##  0.0  22
##  0.5  33
##  1.0  44
##  1.5  56
##  2.0  67
##  2.5  78
##  3.0  89
##  3.5 100
##
##
##  Total Points Predicted Value
##           73           0.001
##           85           0.010
##           97           0.100
##          108           0.500
##          120           0.900
##          132           0.990
##          144           0.999

```

Unha vez obtidas as regras de clasificación, aínda que xa se dixo anteriormente que o punto de corte considerado para cada paciente pode variar, é conveniente establecer un punto de corte xeral para tomar como referencia. Isto logrouse mediante técnicas de remostraxe *wild bootstrap* con 1000 mostras *bootstrap*, escollendo para cada mostra o máximo  $c$  de forma que non se tiveran falsos negativos e elixindo finalmente o percentil 5 de todos estes puntos de corte. Segundo esta metodoloxía, os puntos de corte para cada modelo son:

$$c_1 = 0.03, \quad c_2 = 0.04$$

sendo  $c_1$  o punto de corte para o modelo 1 e  $c_2$  o punto de corte para o modelo 2. Entón, para

estes puntos de corte, as puntuacións segundo o nomograma son:

$$puntos_1 = 69, \quad puntos_2 = 92$$

é dicir, se un paciente ten 69 puntos ou máis segundo o modelo 1, será clasificado como positivo, mentres que se se aplica o modelo 2, terá que igualar ou superar os 92 puntos para que iso ocorra.

Para finalizar coa base de datos correspondente a pacientes con derrame pleural, avaliarase o optimismo e o sobreaxuste no referente ás  $AUC$ 's dos modelos. Primeiro calcularase a área baixo a curva corrixida por optimismo mediante metodoloxía *bootstrap*, obténdose así a  $AUC$  corrixida de cada modelo:

$$AUC_c^1 = 0.965, \quad AUC_c^2 = 0.983$$

sendo  $AUC_c^i$  a  $AUC$  corrixida para cada modelo  $i \in \{1, 2\}$ . Se se teñen en conta os intervalos de confianza para as áreas baixo a curva de cada modelo (ver Figura 3.6.2), pódese ver que a diferenza das  $AUC$ 's orixinais e das corrixidas non é significativa, xa que estas últimas están dentro do intervalo de confianza das primeiras.

Por último, utilizáronse as metodoloxías *shrinkage* uniforme e *Lasso* para recalcular os coeficientes dos dous modelos de forma que se corrixe un pouco o sobreaxuste. Os resultados obtidos, con comparación de  $AUC$ 's incluídas, pódense ver nos Cadros 3.6.7 e 3.6.8

Variables	Orixinal	<i>shrinkage</i> uniforme	<i>Lasso</i>
$side = 1$	0.207	0.198	<0.001
$side = 2$	2.455	2.346	1.502
$xRay = 1$	6.120	5.848	4.283
$dyspnea = 1$	-1.666	-1.592	-1.228
$aspect = 1$	3.731	3.565	2.470
$AUC$	0.973	0.973	0.976

Cadro 3.6.7: Coeficientes do modelo 1 e  $AUC$ , tanto os do modelo orixinal, como os modificados mediante *shrinkage* uniforme e *Lasso*.

Se se comparan os as  $AUC$ 's dos modelos 1 e 2 orixinais coas súas variantes mediante *shrinkage* uniforme e *Lasso* (ver Cadros 3.6.7 e 3.6.7), obsérvase que son practicamente iguais, polo que en canto a capacidade discriminativa, o cálculo dos modelos con *shrinkage* non mellora os resultados do modelo orixinal.

Variables	Orixinal	<i>shrinkage</i> uniforme	<i>Lasso</i>
<i>side</i> = 1	2.033	1.949	0.990
<i>side</i> = 2	5.576	5.346	3.651
<i>xRay</i> = 1	11.155	10.695	7.963
<i>dyspnea</i> = 1	-3.584	-3.436	-2.529
<i>aspect</i> = 1	7.929	7.602	5.629
<i>lcea</i>	4.327	4.148	3.127
<i>AUC</i>	0.995	0.995	0.995

Cadro 3.6.8: Coeficientes do modelo 2 e *AUC*, tanto os do modelo orixinal, como os modificados mediante *shrinkage* uniforme e *Lasso*.

### 3.7. Conclusións

Para tratar de discernir se un derrame pleural transudado é de causa maligna ou cancerosa ou non, construíronse dous modelos de diagnóstico, un deles considerando só variables clínicas e o outro considerando ademais un marcador tumoral, o CEA. O modelo máis sinxelo está pensado para centros de máis baixo nivel que non contan con orzamento nin medios suficientes para a realización de probas máis específicas, mentres que o outro modelo está pensado para centros que contan cunha unidade específica de pneumoloxía. De entre as 13 posibles variables clínicas a considerar, os modelos finais só conteñen 4 delas, sento estas as referentes ó lado no que se produciu o derrame, a presenza ou ausencia de elementos suxestivos de malignidade nas probas de imaxe, a presenza ou ausencia de dispnea e o aspecto do líquido.

Despois da selección de variables para os dous modelos, chama a atención que variables que se pensaba que tiñan gran influencia, como pode ser o sexo, a idade ou se o paciente fuma ou non, non entraron non modelo e que ter dispnea sexa un factor protector, é dicir, que padecer deste mal diminúa as probabilidades de que a causa do derrame sexa cancerosa.

Os dous modelos presentan unha gran capacidade discriminativa (ver a Figura 3.6.2) e unha calibración moi boa (ver a Figura 3.6.4), aínda que se poden apreciar diferenzas significativas entre os dous (ver o Cadro 3.6.3 e a Figura 3.6.4). Aínda así, cando se tratou de ver a utilidade clínica dos dous modelos mediante curvas de decisión, para probabilidades inferiores ó 0.05 non se acharon diferenzas significativas entre o modelo simple e o ampliado.

En resumo, os dous modelos avaliados son moi bos para o diagnóstico de pacientes con derrame pleural transudado, sendo recomendable o uso do modelo ampliado (o que inclúe o marcador tumoral) nos casos en que sexa posible.

## Capítulo 4

# Terapia de Resincronización Cardíaca

Neste capítulo vaise estudar a base de datos relativa á Terapia de Resincronización Cardíaca (TRC), que foi explicada no Capítulo 2. Neste caso trátase dun estudo lonxitudinal, é dicir, con seguimento ó longo do tempo, sendo estes estudos os que máis información aportan pero tamén os máis custosos e os de maior complexidade no tocante ó tratamento estatístico. As técnicas estatísticas axeitadas para este tipo de estudos son coñecidas como análise de supervivencia.

### 4.1. Análise de supervivencia

A análise de supervivencia é un conxunto de técnicas mediante as que se pretende describir e cuantificar o tempo ata un evento de interese, como pode ser a aparición de certa enfermidade ou o exitus, que é o caso da base de datos que se está a tratar.

Na análise de supervivencia emprégase o termo “fallo” para definir que sucede o evento de interese. Este “fallo”, en termos médicos, pode ser en realidade un “éxito”, coma no caso da desaparición dunha enfermidade, ou un “fallo” de verdade, coma nos casos descritos no párrafo anterior. O “tempo de supervivencia defínese como o tempo transcorrido ata que ten lugar o evento a estudar, polo que é importante coñecer as datas implicadas no estudo.

En análise de supervivencia, a análise de datos pode ser realizada con técnicas paramétricas, semiparamétricas ou non paramétricas. En epidemioloxía, debido a que normalmente non se dispón dunha distribución teórica axustada ós datos, os métodos máis usados son os semi-paramétricos e os non paramétricos, entre os que destacan a regresión de Cox e o método de Kaplan-Meier (para máis información sobre estes modelos, pódese consultar Therneau e Grambsch (2000)).

#### 4.1.1. Censura

Unha característica clave dos estudos de supervivencia é que os datos de seguimento dos pacientes están, polo xeral, incompletos. Isto débese a que só se coñece o tempo exacto de supervivencia do individuos que tiveron o evento de interese durante o período de seguimento.

Ademais, non todos os pacientes que non experimentaron o evento teñen o mesmo tempo de seguimento, xa que este se puido interromper por outras causas, como por exemplo que o paciente se mudara de cidade e deixara de ir ás consultas. Que non se presente o evento durante o período de observación é un caso de censura, xa que non se sabe a ciencia certa se se experimentou ou non máis tarde e, no caso de ser así, o tempo transcorrido ata que o evento tivese lugar. Só se sabe que durante o tempo observado o suxeito en cuestión non tivo o evento. Debido a isto, a variable resposta, é dicir, o tempo ata o evento, presenta dificultades á hora de levar a cabo o estudo. Ademais, existe outra dificultade, que é o non axuste a unha distribución gaussiana (tendo máis ben unha distribución asimétrica), podendo ser unha posible solución realizar transformacións sobre a variable. Estes son os motivos que xustifican o uso de métodos específicos para o estudo deste tipo de variable resposta, sendo o máis usado a análise de supervivencia anteriormente comentada.

Existen varios tipos de censura (Rizopoulos, 2012):

- **Censura por intervalos:** este tipo de censura tense cando se sabe que o evento tivo lugar dentro dun determinado período de tempo, pero non se coñece o momento exacto. Por exemplo, á para saber se un paciente desenvolve unha determinada enfermidade, realízanse diferentes análises. Cando un paciente ten resultado positivo, sábese que desenvolveu a enfermidade entre a anterior visita e a actual, pero non o momento exacto.
- **Censura pola esquerda:** este tipo de censura implica que só se ten coñecemento dos tempos de supervivencia superiores a un certo tempo. Un exemplo disto ten lugar nos centros de desenvolvemento de habilidades temperás en nenos, nos que se ten interese por determinar o momento en que un neno aprende a desenvolver unha certa tarefa, sendo a idade do neno nese momento a variable de interese. É común que algúns dos nenos sexan capaces xa de desenvolver esa tarefa ó comezo do estudo, polo que non se sabería a idade exacta.
- **Censura pola dereita:** é a máis común e implica que só se coñecen os tempos inferiores a un determinado tempo. Por exemplo, se se quere coñecer o tempo ata a aparición de cancro en determinados pacientes e se realiza un estudo a cinco anos, haberá persoas que ó cabo dese período de tempo non presenten a enfermidade. Deste subconxunto de pacientes, algúns presentarán o evento, pero non se saberá o tempo exacto. En realidade, non se terá en conta a aparición posterior do evento. Para este tipo de censura, dise que unha observación  $Z$  (tempo de vida do suxeito) é censurada pola dereita se se descoñece o valor de  $Z$  pero sábese que é maior que un determinado tempo  $C$  (tempo de censura). Existen varios tipos de censura pola dereita:
  - Censura tipo I: ten lugar cando finaliza o seguimento se que o paciente experimentara o evento. Neste tipo de censura, fíxase unha cota inferior común a todos os individuos.
  - Censura tipo II: ten lugar cando o estudo está deseñado para finalizar cando se produce un número fixo  $r$  de fallos ( $r < n$ ).
  - Censura aleatoria: neste caso, durante o estudo, existe un suceso independente do de interese que provoca a saída dun número  $r$  de individuos.

Sexa  $T$  o tempo de supervivencia e  $C$  o tempo de censura. Supoñendo que  $T$  e  $C$  son independentes ó longo do tempo, os datos observados para cada individuo e para os tres tipos de censura son como segue:

$$Z = \min(T, C) = \begin{cases} T, & \text{para observacións non censuradas} \\ C, & \text{para observacións censuradas} \end{cases}$$

$$\delta = \mathbb{I}(T \leq C) = \begin{cases} 1, & \text{para observacións non censuradas} \\ 0, & \text{para observacións censuradas} \end{cases}$$

$\delta$  é a función indicadora da non censura. Aquí, as respostas son observadas só se se atopan por debaixo dun valor  $C$  (é dicir, se  $\delta = 1$ ). Para o caso de censura tipo I, o tempo de censura  $C$  é constante. Cando se ten censura tipo II,  $C$  é o tempo no que se acadan  $r$  eventos, téndose  $\sum_{i=1}^n \delta_i = r$  con  $n$  o número total de individuos. Para o caso de censura aleatoria, o tempo de censura  $C$  é unha variable aleatoria.

#### 4.1.2. Funcións básicas en análise de supervivencia

Sexa  $T$  a variable aleatoria dos tempos de fallo a estudar, tal e como se definiu anteriormente. Entón, por tratarse de tempos,  $T$  é unha variable continua non negativa. A función principalmente usada para describir a distribución de  $T$  é a chamada función de supervivencia. Esta función indica a probabilidade de que o evento teña lugar despois dun tempo  $t$ , é dicir, a probabilidade de sobrevivir  $t$  tempo. Sexan  $f(t)$  a función de densidade de probabilidade de  $T$  e  $F(t)$  a función de distribución acumulada correspondente, é dicir:

$$F(t) = \mathbb{P}(T \leq t) = \int_0^t f(x) dx$$

Entón, aplicando a notación anterior, a función de supervivencia é:

$$S(t) = \mathbb{P}(T > t) = 1 - F(t) = \int_t^{\infty} f(x) dx \quad (4.1.1)$$

A función de supervivencia é unha función non crecente a medida que  $t$  aumenta, con  $S(0) = 1$  e  $S(\infty) = \lim_{t \rightarrow \infty} S(t) = 0$ .

Outra función que xoga un importante papel na análise de supervivencia é a función de risco (en inglés, *hazard rate*), que se denotará como  $\lambda(t)$ . Esta función ten a seguinte expresión:

$$\lambda(t) = \lim_{\Delta t \rightarrow 0} \frac{\mathbb{P}(t \leq T < t + \Delta t | T \geq t)}{\Delta t} = \frac{f(t)}{S(t)}, \quad t > 0 \quad (4.1.2)$$

A función de risco describe o risco instantáneo para un evento no intervalo  $[t, t + \Delta t)$  supoñendo que o evento non tivo lugar anteriormente. Debido á súa definición,  $\lambda(t)$  tamén é coñecida como función de risco instantáneo. Debido a que a función de densidade de probabilidade tamén se pode expresar  $f(t) = \lambda(t) \prod_{Z_{[i]} < t} [1 - \lambda(Z_{[i]})]$ , utilizando (4.1.2) a función de supervivencia pode ser escrita do seguinte xeito:

$$S(t) = \frac{f(t)}{\lambda(t)} = \prod_{Z_{[i]} < t} [1 - \lambda(Z_{[i]})] \quad (4.1.3)$$

onde  $Z_{[1]} \leq \dots \leq Z_{[n]}$  é a mostra ordenada dos tempos  $Z_1, \dots, Z_n$ .

Da ecuación (4.1.1) derívase que  $f(t) = -dS(t)/dt$ . Tendo en conta isto, a expresión (4.1.2) pode ser descrita como segue:

$$\lambda(t) = -\frac{d}{dt} \log S(t) \quad (4.1.4)$$

Integrando (4.1.4) entre 0 e  $t$  e introducindo a condición de contorno  $S(0) = 1$ , obtense un novo xeito de expresar a función supervivencia (4.1.1) para todos os tempos ata  $t$ :

$$S(t) = \exp \left\{ -\int_0^t \lambda(x) dx \right\} = \exp \{-\Lambda(t)\}$$

onde a función  $\Lambda(t) = \int_0^t \lambda(x) dx = -\log S(t)$  é coñecida como función de risco acumulado e representa os riscos acumulados dende o instante 0 ata o tempo  $t$ .

### 4.1.3. Estimador de Kaplan-Meier

O estimador de Kaplan-Meier é o estimador non paramétrico de máxima verosimilitude da función de supervivencia  $S(t)$ . Este estimador está baseado nos tempos de supervivencia de cada individuo, asumindo que a censura é independente do tempo de supervivencia. Polo tanto, a probabilidade de sobrevivir nun período de tempo pódese calcular considerando pequenas particións ou intervalos dese tempo, que se determinan pola ocorrencia do evento. Esta é a razón de que a probabilidade de supervivencia entre dous eventos non varíe. O estimador de Kaplan-Meier supón ademais que a distribución é discreta no canto de continua, tendo lugar os eventos só nos tempos observados. Tendo en conta todos estes supostos, a función de risco  $\lambda(t)$  pode ser estimada do seguinte xeito:

$$\hat{\lambda}(t) = \frac{\bar{N}(t)}{\bar{Y}(t)} \quad (4.1.5)$$

sendo  $\bar{N}(t)$  o total de fallos e  $\bar{Y}(t)$  o número de suxeitos en risco no tempo  $t$ .

O estimador de Kaplan-Meier para a función de supervivencia  $S(t)$  consiste en aplicar (4.1.3) utilizando como estimación da función de risco  $\lambda(t)$  o estimador visto en (4.1.5). Así tense:

$$\hat{S}_{KM}(t) = \prod_{Z_{[i]} < t} [1 - \hat{\lambda}(Z_{[i]})] = \prod_{Z_{[i]} < t} \left[ 1 - \frac{\bar{N}(Z_{[i]})}{\bar{Y}(Z_{[i]})} \right] \quad (4.1.6)$$

Este estimador pode ser representado graficamente mediante unha función graduada chamada curva de Kaplan-Meier. Nesta curva, como xa se dixo anteriormente, a proporción de supervivencia entre eventos permanece constante. Na Figura 4.1.1 pódese ver un exemplo deste gráfico.

Na Figura 4.1.1 tense a representación gráfica do estimador de Kaplan-Meier para a función de supervivencia para unha submostra da mostra orixinal coa que se vai a traballar. A supervivencia acumulada (eixo vertical) é 1 para o momento 0, o que resulta obvio. A medida que



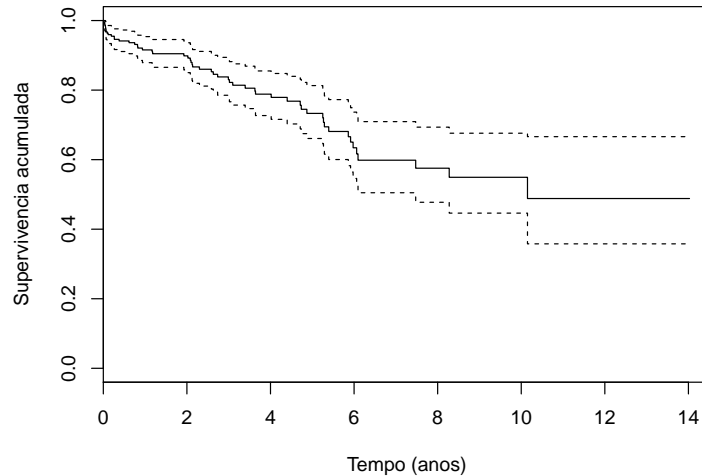


Figura 4.1.1: Exemplo dunha curva de supervivencia de Kaplan-Meier, co intervalo de confianza ó 95 % (liña discontinua).

vai aumentando o tempo (eixo horizontal), a supervivencia acumulada vai caendo de maneira gradual, producíndose un descenso no momento en que se produce algún evento (falecemento), mentres que se mantén constante durante o tempo restante. As liñas discontinuas representan as bandas de confianza ó 95 %.

Ó observar a gráfica, tense unha idea de como evoluciona a mortalidade ó longo do período de estudo. Para este exemplo, tense que ó cabo de dous anos tense unha supervivencia de aproximadamente o 90 %, mentres que ó final do estudo, despois de 14 anos, se ten sobre o 50 % de supervivencia.

En estudos médicos, con asiduidade interesa comparar a supervivencia de dous ou máis grupos de individuos, como poden ser casos ou controis, interesando o efecto do factor que diferencia os grupos na supervivencia. Aínda que non sempre é concluínte, sendo necesario o test log-rank, esta comparación tamén se pode facer representando as curvas de Kaplan-Meier para cada grupo de forma conxunta.

#### 4.1.4. Modelo de regresión de Cox

O modelo de regresión de Cox (1972) é o máis usado para modelizar a relación entre un conxunto de covariables e a supervivencia ou outra variable resposta censurada. Este modelo é a extensión natural do modelo loxístico binario (ver Sección 3.1) para os casos nos que a variable dicotómica a estudo depende do tempo. É dicir, é o máis empregado cando no estudo non só interesa a aparición do evento, senón o momento no que se produce o mesmo.

No modelo de Cox, o risco defínese como segue:

$$\lambda(t|X) = \lambda_0(t)e^{\beta X} \quad (4.1.7)$$

sendo  $X$  o vector de covariables. O modelo presentado en (4.1.7) é un modelo semiparamétrico que consta das seguintes partes:

- $\lambda_0(t)$  constitúe a parte non paramétrica no modelo e é coñecida como función de risco basal. Trátase dun factor de risco común a todos os individuos e non é máis que a función de risco para un tempo  $t$ ,  $\lambda(t)$ , estimada para os valores medios dos preditores.
- A parte paramétrica do modelo, coñecida como puntuación de risco ou *risk score*, é  $e^{\beta X}$ .  $\beta$  é o vector de parámetros da regresión, mentres que  $\beta X$  non é máis que o predictor linear  $\beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p$ .

No modelo loxístico, emprégase o predictor linear do intercepto, mentres que no modelo de regresión de Cox se utiliza a función de risco basal, sendo así a estrutura dos dous modelos moi similar. A cantidade  $e^{\beta_i}$  é a taxa de risco para cada variable, similar á *odds ratio* no modelo loxístico (ver (3.1.1)).

O modelo de Cox tamén se coñece co nome de modelo de riscos proporcionais, xa que o cociente entre os risco para dous suxeitos con vectores de covariables fixo  $X^i$  e  $X^j$  respectivamente é constante no tempo. Isto é:

$$\frac{\lambda(t|X^i)}{\lambda(t|X^j)} = \frac{\lambda_0(t)e^{\beta X^i}}{\lambda_0(t)e^{\beta X^j}} = \frac{e^{\beta X^i}}{e^{\beta X^j}}$$

Pódese dar neste tipo de modelos que as covariables tamén dependan do tempo, resultando agora o modelo:

$$\lambda(t|X(t)) = \lambda_0(t)e^{\beta^t X(t)} \quad (4.1.8)$$

onde  $\beta^t X(t) = \beta_1 X_1(t) + \dots + \beta_p X_p(t)$ .

Para estimar  $\beta$ , emprégase a función de verosimilitude parcial, introducida por Cox (1972). Supóñase que tivo lugar un deceso no tempo  $t^*$ . Entón, a probabilidade de que lle ocorra o evento de interese ó individuo  $i$ -ésimo e non a outro é:

$$L_i(\beta) = \frac{\lambda_0(t^*)r_i(t^*)}{\sum_{j=1}^n Y_j(t^*)\lambda_0(t^*)r_j(t^*)} = \frac{r_i(t^*)}{\sum_{j=1}^n Y_j(t^*)r_j(t^*)}$$

onde  $Y_k(t)$  é a función indicadora de se o individuo  $k$ -ésimo está en risco no tempo  $t$  e  $r_k(t) = \exp\{\beta X^k(t)\}$  é o chamado índice de risco para o suxeito  $k$ -ésimo. Nestas circunstancias, o produto

$$PL(\beta) = \prod_{i=1}^n L_i(\beta)$$

é a xa mencionada verosimilitude parcial. Unha estimación de  $\beta$  pode ser obtida maximizando o logaritmo da verosimilitude parcial ( $\log(L(\beta))$ ) sen necesidade de estimar a función de risco basal  $\lambda_0(t)$ .

O modelo de Cox así definido asume, para variables continuas, que a razón de risco non varía ó cambiar os valores da variable. Por exemplo, se se considera a variable idade, estase asumindo que a razón de risco é a mesma entre os 20 e os 30 anos que entre os 70 e os 80

anos. Como parece obvio por este exemplo, iso non é, polo xeral, verdade. Debido a isto, é necesario abandonar a forma linear de algunhas variables en favor dunha forma funcional para as mesmas.

Aínda que existen varios métodos que permiten darlle unha forma non linear ó efecto das covariables, a que se vai tratar aquí son os *splines* de suavización ou *smoothing splines*. Sexa  $g(x, \beta)$  un *spline* baseado nun número elevado de nós. Entón, escolleranse os coeficientes  $\beta$  para as funcións base de forma que se minimice

$$\theta \sum_{i=1}^n [y_i - g(x_i, \beta)]^2 + (1 - \theta) \int [g''(x, \beta)]^2 dx$$

onde o primeiro termo é a suma de residuos ó cadrado e o segundo termo é a integral do cadrado da segunda derivada da función *spline*  $g(x, \beta)$ . O segundo termo é mínimo para unha recta ( $g''(x) = 0$ ), mentres que é moi grande para unha liña con moita curvatura.  $\theta$  é un parámetro de axuste que se sitúa no intervalo  $[0, 1]$ . Canto máis pequeno sexa  $\theta$ , máis peso terá a curvatura, de forma que a solución óptima converxe á liña de mínimos cadrados (dous graos de liberdade, contando o intercepto). En cambio, se  $\theta \rightarrow 1$ , entón ten máis peso a suma de residuos ó cadrado, converxendo a solución a unha curva de interpolación que pasa por todos os puntos ( $n$  graos de liberdade).

#### 4.1.5. Contrastes de hipóteses

Unha vez axustados os modelos de Kaplan-Meier e de Cox, é preciso comprobar a súa significación. Para poder facer iso, existen tres contrastes diferentes, pero asintoticamente equivalentes.

##### Test de razón de verosimilitude

O test de razón de verosimilitude defínese como segue:

$$LR = 2 \left( \log(L(\hat{\beta})) - \log(L(\beta_0)) \right) \approx \chi_p^2 \quad (4.1.9)$$

sendo  $\hat{\beta}$  a estimación dos parámetros do modelo e  $\beta_0$  os valores iniciais dos coeficientes.

Dos tres contrastes que se ven neste apartado, este é o considerado como máis fiable.

##### Test de Wald

Sexa  $\hat{\beta}$  a estimación de máxima verosimilitude de  $\beta$  e sexa tamén  $I(\beta)$  a matriz de información de Fisher avaliada en  $\beta$ . Entón,  $\hat{\beta}$  é un vector aleatorio con distribución normal de media  $\beta$  e matriz de varianzas covarianzas estimada por  $I^{-1}(\hat{\beta})$ .

O estatístico de contraste é:

$$W = (\hat{\beta} - \beta_0)^t I(\hat{\beta})(\hat{\beta} - \beta_0) \approx \chi_p^2 \quad (4.1.10)$$

Dos tres tests aquí explicados, o test de Wald é considerado como menos fiable.

### Test das puntuacións (*score test*)

Este test é equivalente ó log-rank test mencionado para o caso do estimador de Kaplan-Meier. Esta proba está baseada no vector  $U(\beta) = (U_1(\beta), \dots, U_p(\beta))^t$ , con  $U_i(\beta)$  a derivada parcial da función de verosimilitude parcial con respecto a  $\beta_i$ ,  $i \in \{1, \dots, p\}$ . Cando a hipótese nula é certa (é dicir,  $H_0 : \beta = \beta_0$ ), o vector  $U(\beta)$  ten distribución asintótica normal de medio o vector cero e matriz de varianzas covarianzas  $I(\beta)$ .

O test defínese como:

$$ST = U(\beta_0)^t I(\beta_0)^{-1} U(\beta_0) \approx \chi_p^2$$

Nos tres casos, rexéitase a hipótese nula  $H_0 : \beta = \beta_0$  se o estatístico é maior a  $\chi_{p,\alpha}^2$ , sendo usualmente  $\alpha = 0.95$ .

#### 4.1.6. Análise de residuos

Os residuos de Schoenfeld defínense como unha matriz de  $p$  columnas e cuxas filas son:

$$s_i = X^i(t_i) - \bar{X}(\beta, T_i)$$

onde  $i$  e  $T_i$  son os individuos que experimentan un evento e o tempo de ocorrencia do mesmo respectivamente,  $X^i$  é o vector das covariables para o suxeito  $i$ -ésimo e  $\bar{X}(\beta, T_i)$  é a media das covariables para os individuos que continúan en risco no tempo  $T_i$ . Deste xeito, e seguindo a notación anteriormente empregada, tense:

$$\bar{X}(\beta, T_i) = \frac{\sum_{j=1}^n Y_j(T_i) r_j(T_i) X^j(T_i)}{\sum_{j=1}^n Y_j(T_i) r_j(T_i)}$$

Os residuos de Schoenfeld son útiles para contrastar a proporcionalidade de riscos.

## 4.2. Curva ROC dependente do tempo

Na Sección 3.2 xa se tratou a curva ROC como xeito de comprobar a capacidade diagnóstica dun modelo. Neste caso, o obxectivo é dar unha predición a certo tempo, polo que a curva ROC é insuficiente. Debido a isto, considéranse curvas ROC dependentes do tempo, denotadas como  $ROC(t)$  (véxase Heagerty et al. (2000)). Desta forma, precísanse novas definicións para a sensibilidade e a especificidade dependentes do tempo. As medidas de precisión poden ser usadas para calcular as curvas ROC para un tempo específico.

Debido a que a variable resposta depende do tempo, pódese representar a curva  $ROC(t)$  como *sensibilidade*( $t$ ) fronte a  $1 - \text{especificidade}(t)$ . Existen tres formas de considerar a sensibilidade e a especificidade dependendo de diferentes puntos de vista (Heagerty e Zheng, 2005).

Sexa  $N_i(t) = \mathbb{I}(T_i \leq t)$  a variable indicadora de se o individuo  $i$ -ésimo experimentou o evento ou non, sexa  $M_i$  o marcador a utilizar e sexa  $c$  o punto de corte establecido. Existen diferentes

posibles definicións para os casos e os controis do estudo:

$$\begin{aligned} \text{casos} & : \begin{cases} \text{acumulados } (\mathbb{C}), N(t) = 1, \text{ cando se teñen en conta todos os casos con } T_i \leq t \\ \text{incidentes } (\mathbb{I}), dN(t) = 1, \text{ cando só se teñen en conta os casos con } T_i = t \end{cases} \\ \text{controis} & : \begin{cases} \text{estáticos } (\mathbb{S}), N(t^*) = 0 \\ \text{dinámicos } (\mathbb{D}), N(t) = 0 \end{cases}, \text{ con } t^* \text{ un tempo fixado} \end{aligned}$$

onde  $dN(t) = N(t) - N(t-)$ .

Segundo a forma de combinar os casos e controis descritos, téñense as tres formas de considerar a sensibilidade e a especificidade dependentes do tempo.

- **Acumulada/dinámica:** Usando esta definición tense que a calquera tempo fixado  $t$ , toda a poboación se clasifica como caso ou control baseándose no estado vital nese tempo  $t$ . Por tanto, cada individuo xoga o papel de control cando aínda non experimentou o evento e de caso na situación contraria.

$$\text{sensibilidade}^{\mathbb{C}}(c, t) = \mathbb{P}(M_i > c | T_i \leq t) = \mathbb{P}(M_i > c | N_i(t) = 1)$$

$$\text{especificidade}^{\mathbb{D}}(c, t) = \mathbb{P}(M_i \leq c | T_i > t) = \mathbb{P}(M_i \leq c | N_i(t) = 0)$$

- **Incidente/estática:** Esta definición implica que ningún suxeito cambia o seu estado (sempre caso ou sempre control). Os casos estratificanse segundo o tempo no que ocorre o evento, mentres que os controis son aqueles suxeitos que non teñen o evento durante un período determinado de tempo  $(0, t^*)$ .

$$\text{sensibilidade}^{\mathbb{I}}(c, t) = \mathbb{P}(M_i > c | T_i = t) = \mathbb{P}(M_i > c | dN_i(t) = 1)$$

$$\text{especificidade}^{\mathbb{S}}(c, t) = \mathbb{P}(M_i \leq c | T_i > t^*) = \mathbb{P}(M_i \leq c | N_i(t^*) = 0)$$

- **Incidente/dinámica:** Esta definición supón que un suxeito pode ser control para tempos inferiores ó seu tempo de evento ( $t < T_i$ ) e caso para o tempo de evento ( $t = T_i$ ).

$$\text{sensibilidade}^{\mathbb{I}}(c, t) = \mathbb{P}(M_i > c | T_i = t) = \mathbb{P}(M_i > c | dN_i(t) = 1)$$

$$\text{especificidade}^{\mathbb{D}}(c, t) = \mathbb{P}(M_i \leq c | T_i > t) = \mathbb{P}(M_i \leq c | N_i(t) = 0)$$

Aquí a sensibilidade mide a proporción esperada de suxeitos cun marcador maior ca  $c$  entre aqueles que experimentaron o evento exactamente no tempo  $t$ . A especificidade mide a proporción de individuos cun marcador menor ou igual ó punto de corte  $c$  entre a poboación que aínda non experimentou o evento a tempo  $t$ .

Debido á forma de considerar os casos e os controis, a última definición considérase a máis natural. Ademais, a súa construción permite a extensión para os casos onde as covariables son dependentes do tempo, non tendo máis que cambiar  $M_i$  por  $M_i(t)$ , mentres que isto non é posible para os casos de nos que se considera a sensibilidade acumulada. Estes dous feitos son algúns dos motivos polos que é preferible o uso de sensibilidade incidente e especificidade dinámica, polo que será o método empregado neste traballo.

Como xa se dixo anteriormente, a curva *ROC* non é máis que a representación da sensibilidade fronte a 1 - especificidade. Neste caso, tense  $(\mathbb{I}/\mathbb{D})$  *ROC* para un determinado tempo  $t$  ou  $ROC_t^{\mathbb{I}/\mathbb{D}}$  como a representación da sensibilidade incidente fronte a 1 - especificidade dinámica.

### 4.2.1. Área baixo a curva, $AUC(t)$

Sexa  $p \in [0, 1]$  a especificidade dinámica. Entón a curva  $(\mathbb{I}/\mathbb{D})$   $ROC$  é a función  $ROC_t^{\mathbb{I}/\mathbb{D}}(p)$ , polo que a área baixo a curva,  $AUC(t)$  se pode definir como:

$$AUC(t) = \int_0^1 ROC_t^{\mathbb{I}/\mathbb{D}}(p) dp \quad (4.2.1)$$

A interpretación, para cada tempo  $t$ , é a mesma que para o caso de non haber dependencia do tempo. Canto maior sexa a área baixo a curva, mellor clasifica a proba, mentes que canto máis se aproxime á 0.5, peor discrimina o modelo.

### 4.2.2. Índice de concordancia, $C$

O índice de concordancia  $C$  calcula a probabilidade de que, para un par de suxeitos elixidos ó azar, o suxeito con maior probabilidade de risco experimente o evento antes ca o outro paciente. A área baixo a curva tratada anteriormente mide isto mesmo pero a un certo tempo, mentres que o índice  $C$  se trata dunha medida global.

Existe unha relación moi próxima entre estes dous valores ( $AUC$  e índice  $C$ ). Para entendela mellor, suporase o caso no que as curvas  $ROC$  non dependen do tempo. Sexan dúas observacións independentes  $X_{i,i}$  e  $X_{j,j}$ , onde  $X_i$  e  $X_j$  son os valores do marcador para os pacientes  $i, j$ , para  $i \neq j$  con  $i, j \in \{1, \dots, n\}$ , e  $i$  é o indicador de que un paciente presenta a enfermidade ( $i = 1$ ) ou non ( $i = 0$ ). Entón, por definición, tense

$$C = \mathbb{P}(X_i > X_j | i = 1, j = 0)$$

Por tanto, tense que o índice de concordancia  $C$  é unha medida ponderada da área baixo a curva  $ROC$  para prazos de tempo fixo. En consecuencia, cando a curva  $ROC$  depende do tempo, tense:

$$C = \int_t AUC(t)w(t)dt \quad (4.2.2)$$

onde  $w(t) = 2f(t)S(t)$ .

## 4.3. Resultados para os datos de terapia de resincronización cardíaca

O obxectivo deste apartado é crear un modelo para intentar establecer de que maneira inflúen certas variables na supervivencia de pacientes ós que se lle practicou a terapia de resincronización cardíaca ou TRC. Para cada paciente, as variables a considerar son as expostas no Apartado 2.2.2 do presente escrito, considerándose primeiro un modelo con covariables non dependentes do tempo, para logo considerar outro modelo coas mesmas variables agás  $fe0$ , que será substituída por  $deltafe$  (que depende da anterior), a cal depende do tempo.

### 4.3. RESULTADOS PARA OS DATOS DE TERAPIA DE RESINCRONIZACIÓN CARDÍACA<sup>51</sup>

Antes de empezar a construír o modelo, aplicárase o estimador de Kaplan-Meier ó global dos datos e tamén agrupando segundo as variables discretas, tentando ver se hai unha diferenza significativa entre os diferentes grupos definidos polas variables (ver (4.1.6)).

No Cadro 4.3.1 pódense ver algúns dos resultados do estimador de Kaplan-Meier para os datos de TRC. No cadro aparecen os intervalos de tempo (en anos), o número de suxeitos en risco, o número de eventos sucedidos en cada instante, a probabilidade de supervivencia nos tempos dados, o erro estándar e os intervalos ó 95 %.

Tempo	nº risco	nº eventos	supervivencia	erro estándar	inf. 95 %	sup. 95 %
0.00821	328	1	0.997	0.00304	0.9910	1.000
1.07871	256	1	0.889	0.01775	0.8548	0.924
5.01300	87	1	0.605	0.03324	0.5437	0.674
6.06160	54	1	0.507	0.03826	0.4372	0.588
12.11225	3	1	0.182	0.08437	0.0732	0.451

Cadro 4.3.1: Táboa de vida para pacientes sometidos a terapia de resincronización cardíaca.

No Cadro 4.3.1 pódese apreciar que o primeiro paciente morre transcorridos tan só 3 días dende a intervención. Aproximadamente o 89 % dos pacientes sobreviven despois dun ano de intervención, mentres que este porcentaxe se reduce ó 60 % tras 5 anos. A metade dos pacientes sobreviven 6 anos. Finalmente, aproximadamente o 18 % dos pacientes sobrevive despois de 12 anos. O número de pacientes en risco ó comezo da análise é o número de pacientes,  $n = 328$ . A información extraída do estimador de Kaplan-Meier se pode ver mellor mediante a curva de Kaplan-Meier, a cal se pode ver na Figura 4.3.1.

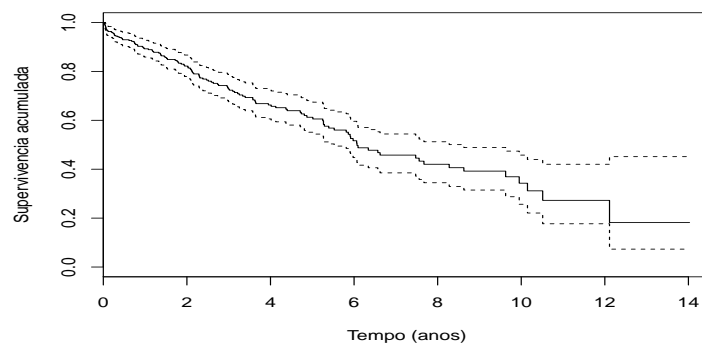


Figura 4.3.1: Curva de supervivencia de Kaplan-Meier, co intervalo de confianza ó 95 % (liña discontinua), para os datos de TRC.

Observando a gráfica da Figura 4.3.1 pódese ver como evoluciona a mortalidade ó longo do período de estudo dunha forma máis sinxela que analizando directamente os resultados do estimador (ver Cadro 4.3.1). Debido a unha máis sinxela interpretación, para o resto de variables a considerar só se amosará a curva de Kaplan-Meier.

Véxase agora a supervivencia segundo a etioloxía. Como xa se vira anteriormente, un 63.72 % dos pacientes teñen miocardiopatía dilatada, mentres que o 36.28 % a ten isquémica. Na parte esquerda Figura 4.3.2 se pode ver a curva de Kaplan-Meier para os dous grupos de pacientes, en vermello para dilatada e en azul para isquémica. Na imaxe parece que os pacientes con miocardiopatía dilatada teñen unha maior supervivencia ca os pacientes con miocardiopatía isquémica. Para verificalo, realízase o test log-rank, obténdose  $p - \text{valor} = 0.001$ , polo que se confirma a suposición.

Na parte dereita da Figura 4.3.2 pódese ver a curva de Kaplan-Meier segundo a clase funcional NYHA. Recórdese que o 24.09 % dos pacientes pertencen á clase II, o 70.12 % á clase III e tan só o 5.79 % pertencen á clase IV. No gráfico pódense apreciar diferenzas entre as clases II e III e a clase IV. Realizando o test log-rank, o  $p - \text{valor}$  obtido é 0.042, polo que se admiten diferenzas significativas entre as tres clases. Aínda así, no é posible establecer con certeza con este test se hai diferenzas entre a supervivencia nos pacientes de clase II e clase III, posiblemente debido á falta de pacientes na primeira categoría.

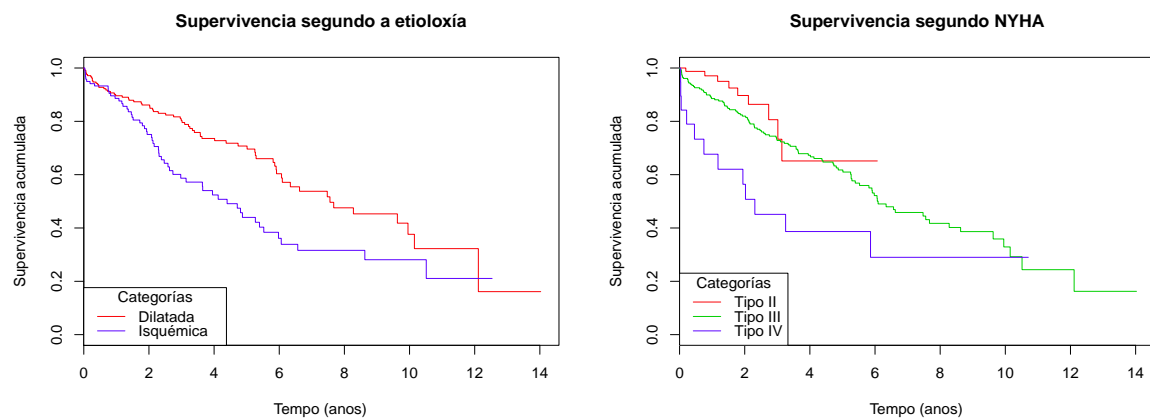


Figura 4.3.2: Esquerda: Curva de supervivencia de Kaplan-Meier para os datos de TRC segundo a etioloxía. Dereita: Curva de supervivencia de Kaplan-Meier para os datos de TRC segundo a clase NYHA.

O seguinte factor a considerar é o ritmo cardíaco no momento do implante. O 62.5 % dos pacientes presentaban ritmo sinusal, mentres que o 37.5 % tiñan fibrilación auricular. Na gráfica esquerda da Figura 4.3.3 pódese apreciar, como era de esperar, que os pacientes con ritmo sinusal no momento do implante presentan unha maior supervivencia ca os pacientes que tiñan fibrilación auricular. Tras realizar o log-rank test, atópanse evidencias significativas deste feito debido a que o  $p - \text{valor}$  é de 0.046.

Se se ten en conta a presenza ou ausencia de bloqueo de rama esquerda (o 60.37 % presentan bloqueo de rama esquerda), na gráfica dereita da Figura 4.3.3 pódese ver que non existen diferenzas á hora da supervivencia, feito que se corrobora ó realizar o log-rank test ( $p -$



### 4.3. RESULTADOS PARA OS DATOS DE TERAPIA DE RESINCRONIZACIÓN CARDÍACA<sup>53</sup>

*valor* = 0.612). Isto sorprende moito xa que, como xa se explicara anteriormente, en estudos previos resultou que a presenza de bloqueo de rama esquerda era un factor protector fronte á mortalidade, é dicir, que os pacientes con bloqueo de rama esquerda sobrevivían máis tempo ca os que non o padecían. Unha das hipóteses que barallan os doutores é o baixo número de pacientes en ditos estudos.

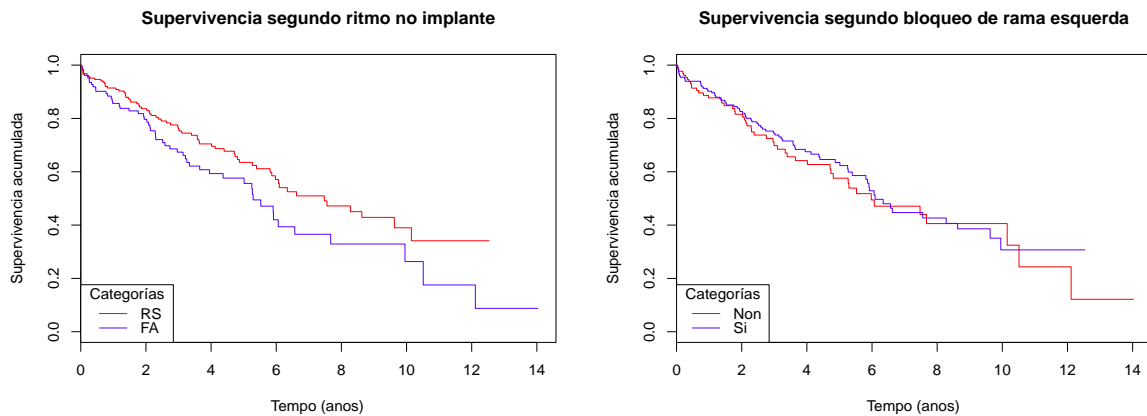


Figura 4.3.3: Curvas de supervivencia de Kaplan-Meier segundo diferentes factores. Esquerda: segundo ritmo no momento do implante. Dereita: segundo presenza de bloqueo de rama esquerda.

Finalmente, pero non menos importante, estúdase o efecto do sexo sobre a supervivencia. Recórdese que o 77.13% dos pacientes son varóns. Na Figura 4.3.4, tense a curva de Kaplan-Meier segundo os sexos, podéndose apreciar unha clara diferenza a favor das mulleres, sendo este feito corroborado polo log-rank test ( $p$ -valor =  $p$  = 0.006).

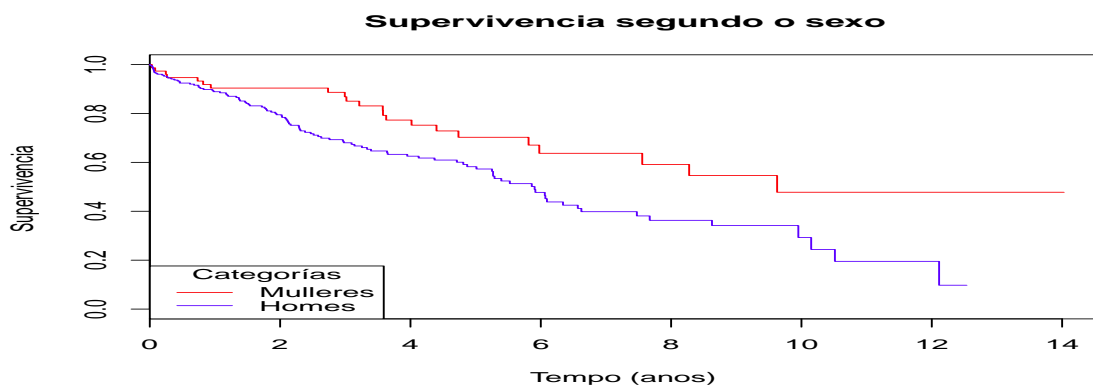


Figura 4.3.4: Curva de supervivencia de Kaplan-Meier segundo o sexo.

Para analizar como inflúen na supervivencia as catro variables continuas, axustarase un modelo de Cox permitindo suavización. Na parte superior esquerda da Figura 4.3.5 tense o efecto da taxa de filtrado glomerular (*tfg*) na mortalidade. Pódese ver que a medida que aumenta a taxa de filtrado glomerular, diminúe a mortalidade (como era de esperar), ata un valor de

aproximadamente  $75 \text{ mg} \cdot \text{ml}/\text{m}^3$ , a partir do cal parece manterse constante.

Os niveis de hemoglobina parecen seguir un comportamento similar (gráfica superior dereita da Figura 4.3.5), de modo que a medida que aumenta a hemoglobina diminúe a mortalidade, ata chegar ós  $14 \text{ g/dL}$ , a partir dos cales parece que aumenta a mortalidade, pero de forma non significativa.

En canto á idade, na gráfica inferior da Figura 4.3.5 pódese ve que a relación entre esta variable e a mortalidade é crecente, linear e significativa.

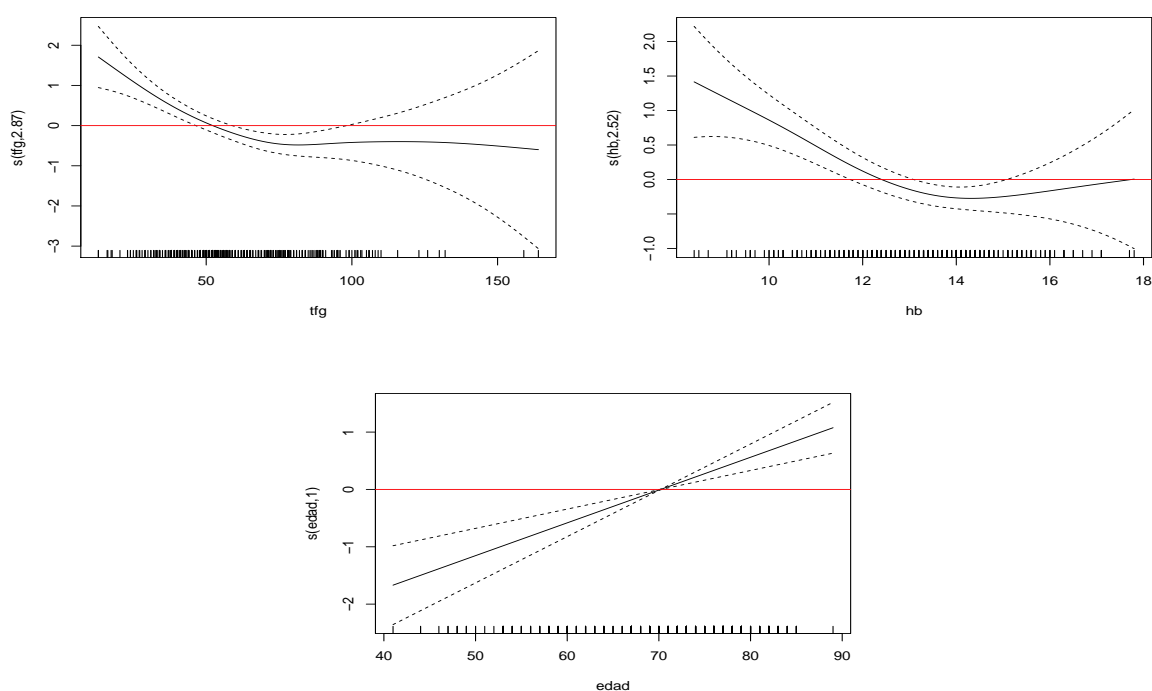


Figura 4.3.5: Efecto das covariables continuas sobre a mortalidade. Arriba esquerda: taxa de filtrado glomerular. Arriba dereita: hemoglobina. Abaixo: idade.

Se se considera a fracción de execución no momento do procedemento,  $fe0$ , na gráfica esquerda da Figura 4.3.6, pódese ver que a relación é linear e decrecente, pero non significativa. Finalmente, considérase a única variable dependente do tempo, a porcentaxe de variación da fracción de execución ou  $\text{deltafe}$ . Na gráfica dereita da Figura 4.3.6 apréciase que a relación entre esta variable e a mortalidade é xusto a oposta á que había coa idade, é dicir, é linear pero decrecente, e tamén significativa.

Unha vez visto o efecto de cada variable por separado na supervivencia, axustaranse dous modelos de Cox (ver (4.1.8)): un con todas as variables basais e outro con unha variable dependente do tempo, os cales chamaremos  $\text{cox1}$  e  $\text{cox2}$  para maior comodidade. No Cadro 4.3.2 tense o axuste correspondente ó modelo  $\text{cox1}$ . No Cadro 4.3.2 obsérvase que as únicas variables significativas do modelo son a idade, o sexo, a fracción de execución basal e a taxa de filtrado glomerular. Vese así que algunhas das variables que eran significativas na análise univariante perderon a súa significación e viceversa. Cabe destacar que a variable  $fe0$  ten un

### 4.3. RESULTADOS PARA OS DATOS DE TERAPIA DE RESINCRONIZACIÓN CARDÍACA55

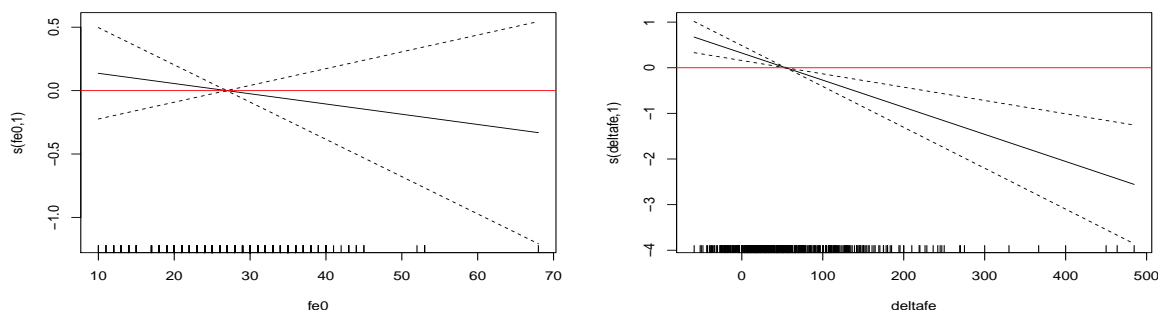


Figura 4.3.6: Efecto das covariables continuas sobre a mortalidade. Esquerda: fracción de exección previa ó implante. Dereita: porcentaxe de variación da fracción de exección.

Variable	Coefficiente	Erro estándar	Graos de liberdade	$p$ – valor
<i>etiolo</i> = Isquémico	0.319	0.201	-	0.113
<i>nyha</i> = Clase III	0.668	0.363	-	0.065
<i>nyha</i> = Clase IV	0.823	0.467	-	0.078
<i>hb</i>	-0.084	0.058	-	0.144
<i>ritmoimpl</i> = FA	0.174	0.195	-	0.372
<i>BRI</i> = Si	0.092	0.195	-	0.639
<i>edad</i>	0.048	0.014	-	<0.001
<i>sexo</i> = Varón	0.698	0.272	-	0.010
<i>fe0</i>	-0.029	0.013	-	0.030
<i>s(tfg)</i>	-	-	3.162	0.003

Cadro 4.3.2: Variables incluídas no modelo *cox1*, xunto coa súa estimación, o seu erro estándar, o seu  $p$  – valor e os seus graos de liberdade efectivos no caso das variables suavizadas.

coeficiente negativo, de forma que a mortalidade diminúe a medida que aumenta esta variable (como era de esperar). Como xa se vira na análise univariante, os homes presentan unha maior mortalidade e, a medida que aumenta a idade, diminúe a supervivencia. Na Figura 4.3.7 pódese ver o efecto da variable suavizada do modelo *cox1*. Tanto os test de razón de verosimilitudes (ver (4.1.9)) como o test de Wald (ver 4.1.10)) dan un  $p$  – valor menor a 0.001, do que se extrae que o modelo é significativo. Ademais, o contraste baseado nos residuos de Schoenfeld ten como  $p$  – valor 0.919, polo que non hai evidencias para rexeitar a hipótese nula de proporcionalidade de riscos.

Véxase agora o que ocorre co modelo *cox2*. No Cadro 4.3.3 tense o axuste correspondente a dito modelo.

Os resultados do modelo *cox2* amosan as mesmas variables significativas, coa inclusión dos

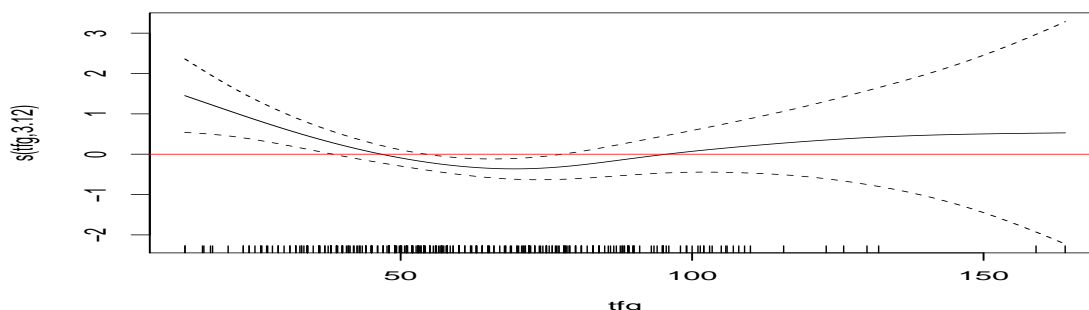


Figura 4.3.7: Efecto da taxa de filtrado glomerular sobre a mortalidade no modelo *cox1*.

Variable	Coefficiente	Erro estándar	Graos de liberdade	<i>p</i> – valor
<i>etiolo</i> = Isquémico	0.299	0.179	-	0.094
<i>nyha</i> = Clase III	0.454	0.334	-	0.175
<i>nyha</i> = Clase IV	0.494	0.440	-	0.262
<i>ritmoimpl</i> = FA	0.203	0.177	-	0.250
<i>BRI</i> = Si	0.179	0.179	-	0.317
<i>sexo</i> = Varón	0.645	0.260	-	0.013
<i>deltafe</i>	-0.004	0.002	-	0.022
<i>s(tfg)</i>	-	-	4.146	<0.001
<i>s(hb)</i>	-	-	2.813	0.026
<i>s(edad)</i>	-	-	1.834	<0.001

Cadro 4.3.3: Variables incluídas no modelo *cox2*, xunto coa súa estimación, o seu erro estándar, o seu *p* – valor e os seus graos de liberdade efectivos no caso das variables suavizadas.

niveis de hemoglobina, *hb*. Ademais, neste modelo as variables *edad* e *hb* son non lineares, a diferenza do que pasaba no modelo *cox1*. En canto a interpretación dos coeficientes, os coeficientes das variables significativas interprétanse dun xeito similar ó visto para o modelo *cox1*.

Na Figura 4.3.8 tense a representación gráfica dos efectos das variables suavizadas do modelo *cox2*.

Da Figura 4.3.8 extráese que tanto co aumento na taxa de filtrado glomerular como nos niveis de hemoglobina diminúe a mortalidade ata certo valor, no que volve aumentar. Ademais, tamén se pode ver que a idade é practicamente linear.

O *p* – valor do test baseado nos residuos de Schoenfeld é 0.995, polo que non hai evidencias

### 4.3. RESULTADOS PARA OS DATOS DE TERAPIA DE RESINCRONIZACIÓN CARDÍACA57

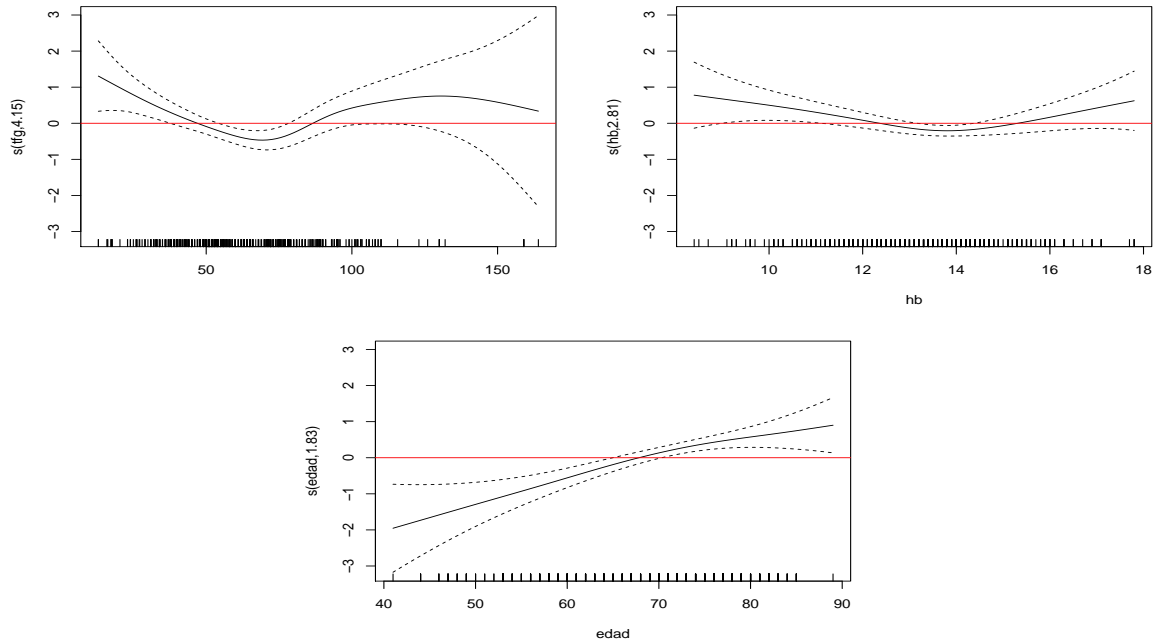


Figura 4.3.8: Efecto das covariables continuas sobre a mortalidade segundo o modelo *cox2*. Arriba esquerda: taxa de filtrado glomerular. Arriba dereita: hemoglobina. Abaixo: idade.

para rexeitar a hipótese de proporcionalidade de riscos. O  $p$ -valor correspondente ó test de razón de verosimilitudes, así como o correspondente ó test de Wald, é menor ca 0.001, polo que se pode considerar o modelo como significativo.

A continuación comparárase o poder discriminativo dos dous modelos mediante  $ROC_t^{I/D}$ . Na Figura 4.3.9 tense a representación gráfica da curva (I/D)  $ROC$  para tempos 0.5 (negro), 1 (vermello), 5 (verde) e 10 anos (azul).

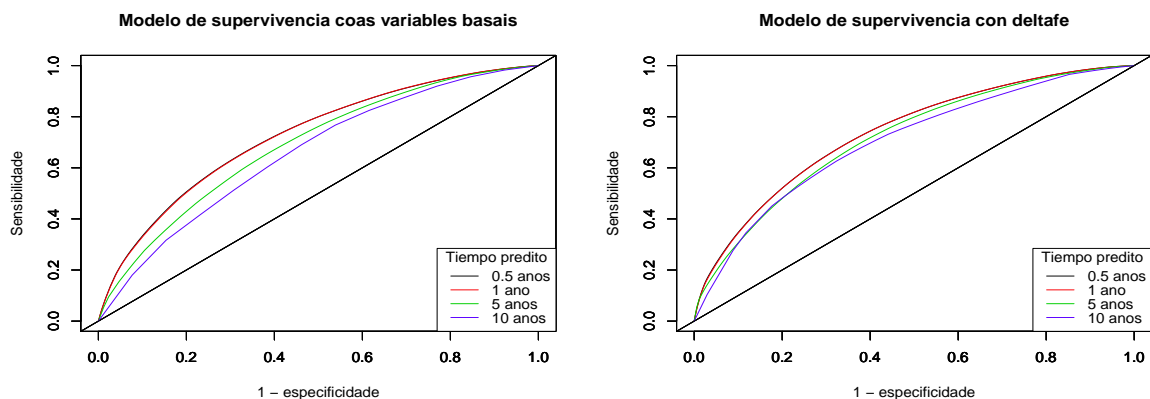


Figura 4.3.9: Curva  $ROC_t^{I/D}$  para tempos  $t \in \{0.5, 1, 5, 10\}$ . Esquerda: modelo *cox1*. Dereita: modelo *cox2*.

Na Figura 4.3.9 pódese apreciar que ambos modelos discriminan de forma moi parecida. Destaca o feito de que o modelo *cox2* sexa capaz de manter mellor a súa capacidade discriminativa,

aínda que non é sorprendente debido a que este modelo contén unha variable dependente do tempo, cousa que no pasa co modelo *cox1*. No cadro 4.3.4 atópanse as *AUC* dos dous modelos para cada un dos tempos indicados (ver (4.2.1)).

Modelo	0.5 anos	1 ano	5 anos	10 anos
<i>cox1</i>	0.725	0.724	0.689	0.657
<i>cox2</i>	0.737	0.737	0.718	0.702

Cadro 4.3.4: Área baixo a curva para os modelos *cox1* e *cox2* para  $t \in \{0.5, 1, 5, 10\}$ .

Na gráfica da esquerda da Figura 4.3.10 tense a representación gráfica das *AUC*'s dos dous modelos ó longo do tempo ata  $t = 10$  anos. Como xa se deducira antes, o modelo *cox2* presenta maior área baixo a curva, polo que discrimina mellor. Para ver se as *AUC*'s dos dous modelos ó longo do tempo se poden considerar distintas, estimáronse as bandas de confianza ó 95% mediante metodoloxía *bootstrap* para 500 mostras *bootstrap* da diferenza de *AUC*'s. Se o 0 cae dentro das bandas de confianza, pódese considerar que non hai diferenzas significativas

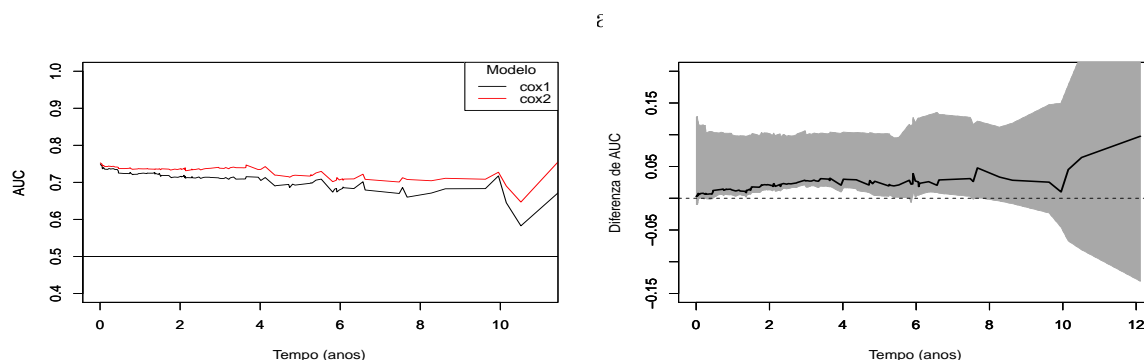


Figura 4.3.10: Esquerda: *AUC*'s ó longo de 10 anos para os modelos *cox1* (negro) e *cox2* (vermello). Esquerda: representación gráfica da diferenza de *AUC*'s para os modelos *cox1* e *cox2*, así como as súas bandas de confianza ó 95%.

Na gráfica da dereita da Figura 4.3.10 vese que a diferenza entre as áreas baixo a curva dos dous modelos é significativa practicamente ata os 8 anos. Aínda que o modelo *cox2* parece ser mellor, é máis interesante o modelo *cox1*, xa que só inclúe variables basais, sendo así posible aplicalo a cada paciente antes de recibir a terapia de resincronización cardíaca. Cabe destacar que a *AUC* dos dous modelos se move en torno a 0.7, xa que outros modelos ou criterios aplicados ós casos de enfermidades cardíacas, como son por exemplo os *scores* de risco cardiovascular, se moven tamén en torno a ese valor. A modo ilustrativo, na Figura 4.3.11 móstrase a curva (I/D) *ROC* para  $t = 1$  ano coas súas bandas de confianza ó 95%. Estas bandas foron estimadas por metodoloxía *bootstrap* utilizando 1000 mostras *bootstrap*.

Finalmente, o índice de concordancia  $C$  proporciona unha medida global con respecto á *AUC*

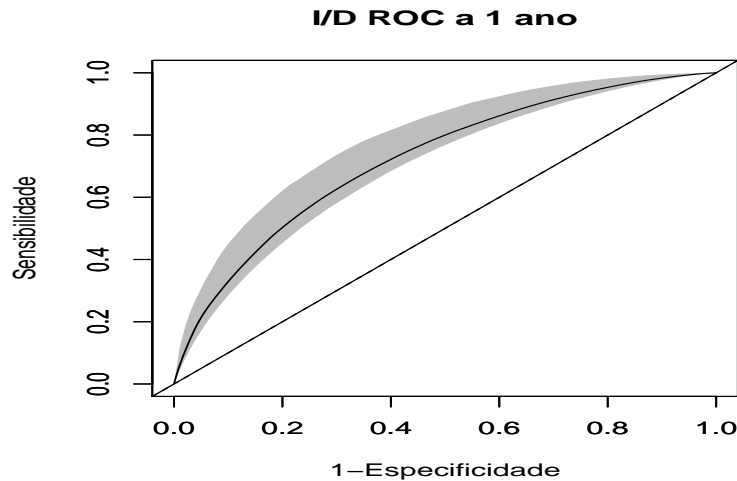


Figura 4.3.11: Representación da curva (I/D) *ROC* para  $t = 1$  ano coas súas bandas de confianza ó 95 %.

(ver (4.2.2)). Para cada un dos modelos, o índice de concordancia

$$C_{cox1} = 0.73 \quad C_{cox2} = 0.741$$

sendo, como cabía esperar, maior o índice para o modelo *cox2*, onde se considera unha variable dependente do tempo.

## 4.4. Conclusións

Para tratar de ver que aspectos inflúen na supervivencia de pacientes con insuficiencia cardíaca sometidos á terapia de resincronización cardíaca ou TRC, construíronse dous modelos de supervivencia de Cox, un que só conta con variables basais e outro que conta con unha variable dependente do tempo.

Despois do axuste dos dous modelos, destaca o feito de que o a presenza de bloqueo de rama esquerda non sexa significativo, xa que en estudos previos se determinou que este era un factor protector. Pero non só a falta de significación do bloqueo de rama esquerda é reseñable, senón a ausencia de significación da maioría das variables clínicas consideradas, como poden ser o índice de clasificación *nyha* ou o ritmo cardíaco no momento do implante. Como cabía esperar, a maior idade maior é a mortalidade, oposto ó que ocorre coa fracción de exección (tanto considerando a fracción de exección basal como a porcentaxe de variación da fracción de exección ó longo do tempo), algo tamén predicible, xa que, como se explicou na introdución, a fracción de exección é un indicador da función cardíaca, sendo esta mellor canto maior é a fracción de exección.

Os dous modelos discriminan en torno a un 70% (ver gráfico esquerdo da Figura 4.3.10), que coincide de forma aproximada á discriminación obtida en outros estudos relacionados con

afeccións cardíacas. Aínda así, e como é esperable, discrimina dun mellor xeito o modelo *cox2*, que inclúe unha variable dependente do tempo (ver gráfico dereito da Figura 4.3.10), xa que son medicións que se van actualizando a medida que pasan os anos, polo que as predicións a longo prazo son máis precisas.

Aínda que é visiblemente mellor o modelo *cox2*, o modelo *cox1* é moi interesante na práctica debido a que só contén variables basais, polo que serve como predictor da supervivencia dun paciente no momento en que chega á unidade de cardioloxía, antes de realizarlle o procedemento.

Por outra banda, o modelo *cox2* é unha primeira aproximación xa non para predición, senón para estimar o comportamento da supervivencia dos pacientes ó longo do tempo. Pero tendo en conta a presenza de covariables dependentes do tempo, un xeito máis preciso (e máis recente) de tratar supervivencia con datos lonxitudinais son os chamados *joint models*, que combinan análise de supervivencia mediante modelos de Cox con análise de datos lonxitudinais mediante modelos mixtos (Rizopoulos, 2012). Esta é a liña de traballo coa que se continuará a análise desta base de datos na unidade de epidemioloxía do Hospital Clínico Universitario de Santiago de Compostela.



## Capítulo 5

# Prácticas realizadas da Unidade de Epidemioloxía

Este capítulo final consiste nun breve resumo das prácticas realizadas na unidade de epidemioloxía clínica do Hospital Clínico Universitario de Santiago de Compostela. Ata o de agora realizouse a descrición dun caso completo de forma máis pormenorizada.

Entre as tarefas realizadas durante as prácticas, atópanse outras consultas ás que se lle tratou de dar resposta. Algunhas delas son:

- Evolución do tamaño dun *Schwannoma*, que é un tipo de tumor que se localiza na cavidade auricular. A base conta con tan só 81 pacientes diagnosticados que foron sometidos a radiocirurxía para tentar reducir o tumor. Este procedemento é menos invasivo que os procedementos aplicados ata o momento, sendo o obxectivo analizar, mediante modelos mixtos, a redución no tamaño do tumor para, en último momento, comparalo cos resultados obtidos mediante os outros procedementos.
- Un estudo da influencia da estación do ano e da dieta nos niveis de vitamina D en poboación infantil. Para iso con 146 nenos ós que se lles mediron os niveis de vitamina D en dous momentos separados 6 meses no tempo. Para isto foi necesario o axuste de modelos mixtos co paquete `gamlss`, xa que os datos non seguían unha distribución das consideradas “usuais”. Ademais, para o caso da dieta foi necesario un análise de compoñentes principais previo para reducir o número de variables.
- Determinación dos valores de referencia para o procedemento *vHIT* de detección da vertixe. Cunha base de datos moi pequena (40 individuos), divididos en 5 homes e 5 mulleres dentro das seguintes franxas de idade: [0, 20], [21, 40], [41, 60], [61, 80]. Para cada un destes suxeitos están recollidos, entre outros datos, o sexo, a idade e as ganancias obtidas para cada unha das 6 canles na proba *vHIT*. O *vHIT* ou proba de impulso cefálico cuantificada, permite realizar proba clínica do impulso cefálico mediante unha micro cámara de vídeo de alta velocidade e coa axuda de xiroscopios integrados e un sistema de rexistro comparar a velocidade do movemento ocular co da cabeza. Isto permite cuantificar a relación entre ambos movementos detectando diferenzas anormais e, ademais, rexistrar

movimentos oculares anómalos que aparecen en pacientes con lesións vestibulares. Para poder facer isto, foi necesario de novo o uso do paquete **gamlss**, axustando primeiro as distribucións de cada canle e despois calculando os valores de referencia de cada unha tendo en conta o sexo e a idade.

- Caracterización dos perfís de glicosa en poboación non diabética. O obxectivo neste caso era dar unha visión exploratoria das curvas de glicosa de 581, así como atopar individuos pre-diabéticos. As curvas de glicosa foron recollidas mediante monitorización continua da glicosa nos pacientes, téndose así os niveis de glicosa durante as 24 horas do día, durante 6 días cada 5 minutos. Para isto foi necesario a análise de datos funcionais.

Pero non só se tratou de responder a consultas formuladas polos facultativos, senón que tamén se realizaron sesións informativas sobre a elaboración de proxectos de investigación e os ensaios clínicos.

### **5.0.1. Valoración das prácticas**

A realización das prácticas na unidade de epidemioloxía clínica do Hospital Clínico Universitario de Santiago de Compostela supón non só o perfeccionamento das técnicas vistas no mestrado, senón a aprendizaxe de nova metodoloxía. Ademais, o aspecto máis importante é a aplicación de dita metodoloxía a problemas reais como nunca se fixo no mestrado, tendo incluso que realizar un proceso previo de preparación de bases de datos, xa que en moitos casos non chegan á unidade dunha forma axeitada para a súa análise.

Estas prácticas supoñen tamén unha grande oportunidade para tratar cara a cara cos usuarios, coa dificultade que iso entraña, sobre todo tendo en conta que moitos deles entenden da súa materia pero non de estatística. Este último é un problema moi común visto na unidade de epidemioloxía e que supón un gran esforzo por parte do médico e do estatístico para poder chegar a un entendemento.

Tamén son moi útiles para obter unha visión máis xeral á hora de realizar un estudo. Así como é importante saber resolver os problemas e realizar a análise estatística, é igual de importante ou incluso máis a planificación do estudo (saber que problemas resolver). Unha boa planificación do estudo desde o comezo facilita moito a súa realización.

En consecuencia, estas prácticas supoñen o complemento ideal para a formación académica obtida no Mestrado de Técnicas Estatísticas, tanto pola aprendizaxe metodolóxica como pola experiencia adquirida.

## Apéndice A

# Software e código empregado

Neste anexo móstranse algúns paquetes de R que foron empregados durante a realización do traballo e que resultaron de especial utilidade.

### A.1. Software empregado para analizar os datos de pacientes con derrame pleural

Para a análise da base de datos relativa a pacientes con derrame pleural transudativo, empregáronse múltiples paquetes, como o paquete MASS para a función `stepAIC` utilizada para a selección das variables, a librería `mgcv` coa que se poden axustar modelos lineares xerais (neste caso, un modelo loxístico) ou o paquete `pROC`, co que se calcularon as curvas *ROC* e as correspondentes *AUC*'s dos modelos.

Entre as librerías empregadas, cabe destacar o paquete `rms` de R (Harrell Jr 2016), pensado exclusivamente para o axuste de modelos de regresión loxística. Este paquete ten funcións máis específicas dos modelos loxísticos, como o cálculo da *odds ratio*, xunto cos gráficos utilizados durante este escrito. Tanto as *odds ratio*'s como moitas medidas de axuste xeral e de capacidade de discriminación ( $R^2$  de Nagelkerke, o coeficiente de Brier, *AUC*, etc) son proporcionados polo comando `summary` do modelo. Deste paquete caben destacar dúas funcións:

- `calibrate`: Esta función realiza a calibración do modelo mediante remostraxe *bootstrap*.
- `nomogram`: Esta función calcula o nomograma do modelo, téndose así tanto o gráfico do nomograma como a táboa resumo con todas as puntuacións. Desta forma, proporciónase unha forma cómoda e fácil de entender para a aplicación do modelo predictivo.

Para a realización deste traballo tamén se construíu código propio, tanto para os casos nos que era precisa remostraxe como para algúns cálculos.

```
# Lectura de datos
library(foreign)
```

```

data <- read.spss('dpmimp1.sav', use.value.labels = TRUE, to.data.frame = T)

library(rms)
library(mgcv)

d <- datadist(data)
options(datadist = "d")
mod1 <- lrm(mpe ~ side + xRay + dyspnea + aspect, data = data, x = TRUE,
           y = TRUE)
summary(mod1, side = "0", dyspnea = "0")
plot(summary(mod1, side = "0", dyspnea = "0"), log = T,
     main = "Odds Ratio do modelo 1")
# Odds ratio e intervalos de confianza

mod1g <- glm(mpe ~ side + xRay + dyspnea + aspect, data = data,
            family = "binomial")

mod2 <- lrm(mpe ~ side + xRay + dyspnea + aspect + lcea, data = data,
           x = TRUE, y = TRUE)
mod2
plot(summary(mod2, side = "0", dyspnea = "0"), log = T,
     main = "Odds Ratio do modelo 2")
# Odds ratio e intervalos de confianza

mod2g <- glm(mpe ~ side + xRay + dyspnea + aspect + lcea, data = data,
            family = "binomial")

#####
##### Curvas ROC #####
#####

library(pROC)
### Para os sintomas clínicos
roc1 <- roc(data$mpe , mod1$linear.predictors, ci = TRUE, ci.type = "shape")

### Para as variables analíticas + lcea
roc2 <- roc(data$mpe , mod2$linear.predictors, ci = TRUE, ci.type = "shape")

#####

### Comparación AUC's

roc.test(roc1, roc2, reuse.auc = FALSE)

difauc <- roc2$auc - roc1$auc

```

```

bootstrap.sample <- function(data) {
  res <- data[sample(nrow(data), replace = TRUE), ]
  res
}

B = 1000
AUC1.a <- rep(0, l = B)
AUC2.b <- rep(0, l = B)
AUC <- rep(0, l = B)
set.seed(12345)
for(i in 1:B) {
  data.a <- bootstrap.sample(data)
  mod.a <- glm(mpe ~ side + xRay + dyspnea + aspect, data = data.a,
              family = "binomial")
  mod.b <- glm(mpe ~ side + xRay + dyspnea + aspect +lcea,
              data = data.a, family = "binomial")

  # Creamos os grupos para o cálculo da ROC
  data.a$mpe <- as.numeric(data.a$mpe - 1)

  data.a[, c("p1")] <- mod.a$fitted.values
  data.a[, c("p2")] <- mod.b$fitted.values
  AUC1.a[i] <- roc(data.a$mpe, data.a$p1, ci = TRUE)$auc[1]
  AUC2.b[i] <- roc(data.a$mpe, data.a$p2, ci = TRUE)$auc[1]
}
AUC <- AUC2.b - AUC1.a
ci <- quantile(AUC, probs = c(0.025, 0.975)); ci

#####
##### Calibración #####
#####

set.seed(12345)
cal1 <- calibrate(mod1, B = 300, method = c("boot"))
plot(cal1, main = "Modelo 1", xlab = "Probabilidade predita",
     ylab = "Frecuencia observada", legend = FALSE, subtitles = FALSE,
     conf.int = TRUE, cex.subtitles = .45, riskdist = TRUE,
     scat1d.opts = list(nhistSpike = 200))

set.seed(12345)
cal2 <- calibrate(mod2, B = 300, method = c("boot"))
plot(cal2, main = "Modelo 2", xlab = "Probabilidade predita",
     ylab = "Frecuencia observada", legend = FALSE, subtitles = FALSE,
     conf.int = TRUE, cex.subtitles = .75, riskdist = TRUE,
     scat1d.opts = list(nhistSpike = 200))

```

```
#####
##### NRI #####
#####

nri2 <- function(p = 0.05, outcome, fitted1, fitted2){
  # outcome ten que ser unha variable numérica de 0 e 1
  event <- mean(outcome)

  c1 <- ifelse(fitted1 <= p, 0, 1)
  c2 <- ifelse(fitted2 <= p, 0, 1)

  up <- ifelse(c1 == 0 & c2 == 1, 1, 0)
  down <- ifelse(c1 == 1 & c2 == 0, 1, 0)
  pup <- mean(up)
  pdown <- mean(down)

  eup <- ifelse(data$mpe[up == 1] == 1, 1, 0)
  peup <- mean(eup)
  edown <- ifelse(data$mpe[down == 1] == 1, 1, 0)
  pedown <- mean(edown)

  nri <- ((peup-event)*pup+(event-pedown)*pdown)/(event*(1-event))
  nri
}
nri2(p = 0.05, data$mpe, mod1g$fitted.values, mod2g$fitted.values)

#####
##### Curva de decisión #####
#####

n <- length(data$mpe); n # 281
fitted <- predict(mod1 ,type = "fitted")
fitted2 <- predict(mod2,type = "fitted")
TPCall <- table(data$mpe)[2] / n
FPCall <- table(data$mpe)[1] / n
p <- seq(0,0.20,by = 0.0001)

NBall <- rep(0, length = length(p))
NBmod1 <- rep(0, length = length(p))
NBmod2 <- rep(0, length = length(p))
fp1 <- rep(0, length = length(p))
fn1 <- rep(0, length = length(p))
fp2 <- rep(0, length = length(p))
fn2 <- rep(0, length = length(p))
p1 <- rep(0, length = length(p))
```

```

n1 <- rep(0, length = length(p))
p2 <- rep(0, length = length(p))
n2 <- rep(0, length = length(p))

for (i in 1:length(p)){
  rate <- p[i]/(1 - p[i])
  NBall[i] <- TPCall - FPCall * rate
  f <- ifelse(fitted < p[i], 0, 1)
  if(sum(f) == 0){NBmod1[i] <- 0} else {
    if(sum(f)== n ){NBmod1[i] <- NBall[i]} else {
      tabla <- table(data$mpe, f)
      NBmod1[i] <- tabla[2, 2] / n - tabla[1, 2] / n * rate
      fp1[i] <- tabla[1, 2]
      fn1[i] <- tabla[2, 1]
      p1[i] <- tabla[2, 2] + tabla[1, 2]
      n1[i] <- tabla[1, 1] + tabla[2, 1]}}
  f2 <- ifelse(fitted2 < p[i], 0, 1)
  if(sum(f2) == 0){NBmod2[i] <- 0} else {
    if(sum(f2) == n){NBmod2[i] <- NBall[i]} else {
      tabla2 <- table(data$mpe, f2)
      NBmod2[i] <- tabla2[2,2] / n - tabla2[1, 2] / n * rate
      fp2[i] <- tabla2[1, 2]
      p2[i] <- tabla2[2, 2] + tabla2[1, 2]
      fn2[i] <- tabla2[2, 1]
      n2[i] <- tabla2[1, 1] + tabla2[2, 1]}}
}

par(mfrow = c(1, 1))
plot(p * 100, NBmod2, type = "l", main = "", xlab = "Punto de corte (%)",
      ylab = "Beneficio neto", lwd = 2, ylim = c(-0.1, 0.1), col = 1,
      xlim = c(0, 20), lty = "dashed")
abline(h = 0, lwd = 2, col = "grey")
par(new = TRUE)
plot(p * 100, NBmod1, type = "l", main = "", xlab = "", ylab = "", lwd = 2,
      col = 1, ylim = c(-0.1, 0.1), xlim = c(0, 20), lty = "dotted")
par(new = TRUE)
plot(p * 100, NBall, type = "l", main = "", xlab = "", ylab = "", lwd = 2,
      col = 1, ylim = c(-0.1, 0.1), xlim = c(0, 20))

legend("bottomleft", lwd = 2,
      c("Ningún positivo", "Todos positivos", "Modelo 1", "Modelo 2"),
      col = c("grey", 1, 1, 1), lty = c(1, 1, 3, 2))

```

```

### Bandas de confianza da diferenza ###

difNB <- NBmod2 - NBmod1
dNB1 <- NBmod1 - NBall
dNB2 <- NBmod2 - NBall

set.seed(12345)
B = 1000
n <- length(data$mpe)
fit1 <- rep(0, l = B)
fit2 <- rep(0, l = B)
NB <- matrix(rep(0, length(p) * B), nrow = B)
NB1 <- matrix(rep(0, length(p) * B), nrow = B)
NB2 <- matrix(rep(0, length(p) * B), nrow = B)

for(j in 1:B) {
  data.a <- bootstrap.sample(data)
  mod.1 <- glm(mpe ~ side + xRay + dyspnea + aspect, data = data.a,
              family = "binomial")
  mod.2 <- glm(mpe ~ side + xRay + dyspnea + aspect + lcea,
              data = data.a, family = "binomial")
  fitted <- mod.1$fitted.values
  fitted2 <- mod.2$fitted.values

  TPCall <- table(data.a$mpe)[2] / n
  FPCall <- table(data.a$mpe)[1] / n

  NBmod1 <- rep(0, length = length(p))
  NBmod2 <- rep(0, length = length(p))

  for (i in 1:length(p)){
    rate <- p[i] / (1 - p[i])
    NBall <- TPCall - FPCall * rate
    f <- ifelse(fitted < p[i], 0, 1)
    if(sum(f) == 0){NBmod1[i] <- 0} else {
      if(sum(f) == n){NBmod1[i] <- NBall} else {
        tabla <- table(data.a$mpe, f)
        NBmod1[i] <- tabla[2, 2] / n - tabla[1, 2] / n * rate}}
    f2 <- ifelse(fitted2 < p[i], 0, 1)
    if(sum(f2) == 0){NBmod2[i] <- 0} else {
      if(sum(f2) == n){NBmod2[i] <- NBall} else {
        tabla2 <- table(data.a$mpe, f2)
        NBmod2[i] <- tabla2[2, 2] / n - tabla2[1, 2] / n * rate}}
    # Facemos a diferenza das curvas
    NB[j, i] <- NBmod2[i] - NBmod1[i]
    NB1[j, i] <- NBmod1[i] - NBall
  }
}

```



```

      NB2[j, i] <- NBmod2[i] - NBall
    }
  }

ciNB <- apply(NB, 2, quantile, probs = c(0.025, 0.975))
ciNB1 <- apply(NB1, 2, quantile, probs = c(0.025, 0.975))
ciNB2 <- apply(NB2, 2, quantile, probs = c(0.025, 0.975))

plot(p * 100, difNB, lwd = 2, type = "l", ylim = c(-0.01, 0.05), main = "",
      xlab = "Punto de corte (%)", ylab = "Diferenza de beneficio neto",
      xlim = c(0, 20))
par(new = TRUE)
plot(p * 100, ciNB[1, ], lwd = 1, lty = 2, type = "l", ylab = "", xlab = "",
      ylim = c(-0.01, 0.05), xlim = c(0, 20), col = 'darkgray')
par(new = TRUE)
plot(p * 100, ciNB[2, ], lwd = 1, lty = 2, type = "l", ylab = "", xlab = "",
      ylim = c(-0.01, 0.05), xlim = c(0, 20), col = 'darkgray')
cord.x <- c(0, p*100, 20, 100*p[length(p):1])
cord.y <- c(0, ciNB[2, 1:length(p)], ciNB[1, length(p)], ciNB[1, length(p):1])
polygon(cord.x, cord.y, col = 'darkgray', border = "darkgray")
abline(h = 0, lty = "dashed")
lines(p * 100, difNB, lwd = 2, type = "l")

#####
##### Nomogramas #####
#####

d <- datadist(data)
options(datadist = "d")

nom1 <- nomogram(mod1, fun = function(x) 1/(1 + exp(-x)), # or fun=plogis
                fun.at = c(.001, .01, .1, 0.5, .9, .99, .999),
                funlabel = "Risco de cancro")
plot(nom1, xfrac = .45)
title(main = "Nomograma para o modelo 1")
print(nom1)

nom2 <- nomogram(mod2, fun = function(x) 1/(1 + exp(-x)), # or fun=plogis
                fun.at = c(.001, .01, .1, 0.5, .9, .99, .999),
                funlabel = "Risco de cancro")
plot(nom2, xfrac = .45)
title(main = "Nomograma para o modelo 2")
print(nom2)

```

```
#####
##### Comprobar sobreaxuste #####
#####
```

```
### ROC curve - AUC
```

```
set.seed(12345)
B = 500
AUC1.a <- rep(0, l = B)
AUC1 <- rep(0, l = B)
for(i in 1:B) {
  data.a <- bootstrap.sample(data)
  mod.a <- glm(mpe ~ side + xRay + dyspnea + aspect, data = data.a,
              family = "binomial")

  # Creamos os grupos para o cálculo da ROC
  data.a$mpe <- as.numeric(data.a$mpe)

  # Bootstrap sample
  data.a[, c("p1")] <- mod.a$fitted.values
  AUC1.a[i] <- roc(data.a$mpe, data.a$p1, ci = TRUE)$auc[1]

  # Original sample
  p.o <- predict(mod.a, type = "response", newdata = data)
  AUC1[i] <- roc((data$mpe - 1), p.o, ci = TRUE)$auc[1]
}
AUC1.o <- mean(AUC1.a - AUC1)

m.AUC1.c <- roc(data$mpe , fit$linear.predictors, ci = TRUE)$auc[1] - AUC1.o
```

```
#####
```

```
set.seed(12345)
B = 500
AUC2.b <- rep(0, l = B)
AUC2 <- rep(0, l = B)
for(i in 1:B) {
  data.b <- bootstrap.sample(data)
  mod.b <- glm(mpe ~ side + xRay + dyspnea + aspect + lcea,
              data = data.b, x = TRUE, y = TRUE, family = "binomial")

  # Creamos os grupos para o cálculo da ROC
  data.b$mpe <- as.numeric(data.b$mpe - 1)

  # Bootstrap sample
  data.b[, c("p1")] <- mod.b$fitted
```

```

AUC2.b[i] <- roc(data.b$mpe, data.b$p1, ci = TRUE)$auc[1]

# Original sample
p.o2 <- predict(mod.b, type = "response", newdata = data)
AUC2[i] <- roc((data$mpe - 1), p.o2, ci = TRUE)$auc[1]
}
AUC2.o <- mean(AUC2.b - AUC2)

fit3 <- glm((mpe - 1) ~ side + xRay + dyspnea + aspect + lcea, data = data,
           x = TRUE, y = TRUE, family = "binomial")

m.AUC2.c <- roc(data$mpe, fit3$fitted.values, ci = TRUE)$auc[1] - AUC2.o

```

## A.2. Software empregado para analizar os datos de pacientes sometidos a TRC

Para a análise da base de datos relativa a pacientes sometidos á terapia de resincronización cardíaca, a principal librería utilizada é a librería *survival* (Therneau e Lumley, 2009). Este paquete permite facer a análise de supervivencia de Cox e Kaplan-Meier, así como algúns tests, como é o test de Wald. Cabe destacar que para a construción do modelo de Cox tamén se pode empregar o paquete *mgcv*, o que ocasiona menores problemas de converxencia, sobre todo cando se incorporan ó modelo covariables suavizadas.

Para o cálculo das curvas *ROC* dependentes do tempo, o paquete utilizado foi *risksetROC* (Heagerty e Saha-Chaudhuri, 2012), o cal permite calcular a curva *ROC* dependente do tempo, así como o seu *AUC* polo método incidente/dinámico. As súas funcións principais son:

- **risksetROC**: Calcula a curva *ROC* para un determinado tempo  $t$  e executa o gráfico.
- **risksetAUC**: Calcula o valor da *AUC* ó longo do tempo, producindo tamén un gráfico.

Para a análise desta base de datos tamén se construíu código propio, en especial para os casos nos que era preciso a remostraxe.

```

#Lectura de datos
library(foreign)
datos <- read.spss("trc.sav", to.data.frame = TRUE, use.value.labels = TRUE)

library(mgcv)

fittfg1 <- gam(ttot ~ s(tfg), family = cox.ph(), data = datos, weights = exit)
# exit é a variable exitus pero en continuo
summary(fittfg1)

```

```

plot(fittfg1)
abline(h=0, col=2)

#####
##### Supervivencia dependente do tempo #####
#####

# Teco son as variables que indican o tempo transcorrido ata a visita ó médico
# (en días)
datos$teco11 <- datos$teco1/365.25
datos$teco21 <- datos$teco2/365.25
datos$teco31 <- datos$teco3/365.25
datos$teco41 <- datos$teco4/365.25
datos$tiempo <- datos$ttot

dat <- reshape(datos, varying = list(c("fe1", "fe2", "fe3", "fe4"),
                                     c("teco11", "teco21", "teco31", "teco41"),
                                     c("teco21", "teco31", "teco41", "tiempo"),
                                     c("teco1", "teco2", "teco3", "teco4"),
                                     c("teco2", "teco3", "teco4", "dias")),
              v.names = c("fe", "start", "stop", "inicio", "final"), times = 1:4,
              direction = "long")
#start e stop indican o principio e final do intervalo de tempo entre visitas

dat <- dat[order(dat$id),]
dat$deltafe<-(dat$fe-dat$fe0)/dat$fe0*100

# Para obter eliminar as visitas nas que non hai datos
base <- NULL
for (i in 1:dim(dat)[1]) {
  if (is.na(dat$deltafe[i]) == TRUE) {
    return} else {
    base<-rbind(base,dat[i,])
  }
}

for (i in 1:dim(base)[1]) {
  if (is.na(base$stop[i]) == TRUE) {
    base$stop[i] <- dat$stop[((dat$id == base$id[i]) & (dat$time == 4))]
  }
}

for (i in 1:dim(base)[1]) {
  if (is.na(base$final[i]) == TRUE) {
    base$final[i] <- dat$final[((dat$id == base$id[i]) & (dat$time == 4))]
  }
}

```

```

}

# Variable indicadora de se o evento tivo lugar nese intervalo de tempo
base$event <- rep(0,length = dim(base)[1])
for (i in 1:length(base$event)) {
  if (dat$stop[((dat$id == base$id[i]) & (dat$time == 4))] == base$stop[i]) {
    base$event[i] <- as.numeric(base$exitus[i])-1
  }
}

base$id[is.na(base$start)==TRUE] # 260
b <- base[base$id != 260, ]

b.id <- b[!duplicated(b$id), ]

## Modelo sen covariables dependentes do tempo (fe0)

cox1<- gam(ttot ~ etiol + nyha + s(tfg) + hb + ritmoimpl + BRI + edad +
          sexo + fe0, family = cox.ph(), data = b.id, weights = exit)
summary(cox1)

coxfe <- coxph(Surv(ttot, event == 1) ~ etiol + nyha + pspline(tfg, df = 3.162) +
              hb + ritmoimpl + BRI + sexo + edad + fe0, data = b.id, x = T)
cox.zph(coxfe)
summary(coxfe)

## Modelo con covariables dependentes do tempo (deltafe)

cox2<- gam(ttot ~ etiol + nyha + s(tfg) + s(hb) + ritmoimpl + BRI +
          s(edad) + sexo + deltafe, family = cox.ph(), data = b,
          weights = event)
summary(cox2)

plot(cox2)
abline(h=0, col=2)

coxdeltafe <- coxph(Surv(ttot, event == 1) ~ etiol + nyha +
                  pspline(tfg, df = 4.146) + pspline(hb, df = 2.813) +
                  ritmoimpl + BRI + sexo + pspline(edad, df = 1.834) +
                  deltafe, data = b.id, x = T)
cox.zph(coxdeltafe)
summary(coxdeltafe)

##### ROC dependente do tempo #####

library(risksetROC)

```

```

# Para el modelo coas variables basais

risksetROC(b.id$ttot, status = b.id$exit, marker = coxfe0$linear.predictors,
           predict.time = 0.5, xlab = "1 - especificidade",
           ylab = "Sensibilidade",
           main = "Modelo de supervivencia coas variables basais")

# Para el modelo con delatafe

risksetROC(b$ttot, status = b$exit, marker = coxdeltafe$linear.predictors,
           predict.time = 0.5, xlab = "1 - especificidade",
           ylab = "Sensibilidade", main = "Modelo de supervivencia con delatafe")

#####

auc1 <- risksetAUC(b.id$ttot, status = b.id$exit, marker = coxfe0$linear.predictors,
                  predict.time = 0.5, tmax = 10, xlab = "Tempo (anos)")
par(new=TRUE)
auc2 <- risksetAUC(b$ttot, status = b$exit, marker = coxdeltafe$linear.predictors,
                  predict.time = 0.5, tmax = 10, col = 2, xlab = " ")
legend("topright", legend = c("cox1", "cox2"),
       lwd = 1, col = 1:4, title = "Modelo")

auc <- auc2$AUC - auc1$AUC

bootstrap.sample <- function(data) {
  res <- data[sample(nrow(data), replace = TRUE), ]
  res
}

interpolate <- function( x, y, target ){
  ooo <- order(x)
  x <- x[ooo]
  y <- y[ooo]
  n <- length(target)
  m <- length(x)
  iii <- c( 1:length(x) )
  output <- rep( NA, n )
  for( j in 1:n ){
    index.low <- max( iii[ x <= target[j] ] )
    index.high <- min( iii[ target[j] <= x ] )
    if( target[j] > max(x) ){
      index.low <- m
      index.high <- m
    }
  }
}

```

```

    if( target[j] < min(x) ){
      index.low <- 1
      index.high <- 1
    }
    #
    if( index.low == index.high ){
      output[j] <- y[ index.low ]
    }else{
      dx <- ( target[j] - x[index.low] )/( x[index.high]-x[index.low] )
      dy <- ( y[ index.high ] - y [ index.low ] )
      output[j] <- y[ index.low ] + dx * dy
    }
  }
  output
}

B = 500
AUC <- matrix(rep(0, l = length(auc)*B), nrow=B)
set.seed(12345)
for(l in 1:B){
  data.a <- bootstrap.sample(b.id)
  mod.a <- gam(ttot ~ etiol + nyha + s(tfg) + hb + ritmoimpl + BRI + edad +
               sexo + fe0, family = cox.ph(), data = data.a, weights = event)
  data.b <- NULL
  for (i in 1:length(data.a$id)){
    for (j in 1:length(b$id)){
      if (b$id[j] == data.a$id[i]) {data.b <- rbind(data.b, b[j,])}
    }
  }
  mod.b <- gam(ttot ~ etiol + nyha + s(tfg) + s(hb) + ritmoimpl + BRI +
               s(edad) + sexo + deltafe, family = cox.ph(),
               data = data.b, weights = event)

  data.a$eta5 <- mod.a$linear.predictors
  data.b$eta5 <- mod.b$linear.predictors
  #
  out1 <- risksetAUC(data.a$ttot, status = data.a$exit, marker = data.a$eta5,
                    tmax = 10, plot = F)
  #
  out2 <- risksetAUC(data.b$ttot, status = data.b$exit, marker = data.b$eta5,
                    tmax = 10, plot = F)
  #
  difauc <- out2$AUC - out1$AUC
  for(k in 1:length(auc1$utimes) ){
    AUC[l, k] <- interpolate(out1$utimes, difauc, auc1$utimes[k])
  }
}

```

```

}

ci <- apply(AUC, 2, quantile, probs = c(0.025, 0.975))

plot(auc1$utimes, auc, lwd = 2, type = "l", ylim = c(-0.15, 0.20), main = "",
      xlab = "Tempo (anos)", ylab = "Diferenza de AUC",
      xlim = c(0, 12))
par(new = TRUE)
plot(auc1$utimes, ci[1, ], lwd = 1, lty = 2, type = "l", ylab = "", xlab = "",
      ylim = c(-0.15, 0.2), xlim = c(0, 12), col = 'darkgray')
par(new = TRUE)
plot(auc1$utimes, ci[2, ], lwd = 1, lty = 2, type = "l", ylab = "", xlab = "",
      ylim = c(-0.15, 0.2), xlim = c(0, 12), col = 'darkgray')

cord.x <- c(auc1$utimes, auc1$utimes[length(auc1$utimes):1])
cord.y <- c(ci[1,1:length(auc1$utimes)], ci[2, length(auc1$utimes):1])
polygon(cord.x, cord.y, col = 'darkgray', border = "darkgray")
abline(h = 0, lty = "dashed")
lines(auc1$utimes, auc, lwd = 2, type = "l")

#####
#####                               I/D                               #####
#####

TargetTime <- 1

set.seed(12345)
Nsim <- 1000
FP.save <- c(0:100)/100
TP.save <- matrix( NA, Nsim, length(FP.save))

for( j in 1:Nsim ){
  boot.data <- bootstrap.sample(b.id)
  fit <- gam(ttot ~ etiol + nyha + s(tfg) + hb + ritmoimpl + BRI + edad +
             sexo + fe0, family = cox.ph(), data = boot.data,
             weights = exit)
  ##### get then risk score
  boot.data$eta5 <- fit$linear.predictors
  #
  attach( boot.data )
  #
  nobs <- nrow( b.id)
  #
  out <- risksetROC(ttot, status = exit, marker = eta5,
                   predict.time = TargetTime, span = 0.25 * nobs^(-0.2),
                   plot = F)
}

```



```

#
detach()
#
for( k in 1:length(FP.save)){
  TP.save[ j, k ] <- interpolate( out$FP, out$TP, FP.save[k])
}
}

upper <- apply( TP.save, 2, quantile, 0.975, na.rm = T )
lower <- apply( TP.save, 2, quantile, 0.025, na.rm = T )
#
print( cbind( FP.save, lower, upper ) )
#
#####
##### plot estimate and point-wise 95% CI
#####
#

##### get then risk score
#
b.id$eta5 <- coxfe0$linear.predictors
#
attach(b.id)
#
out <- risksetROC(ttot, status = exit, marker = eta5,
                  predict.time = TargetTime, span = 0.25 * nobs^(-0.2),
                  plot = F)

detach()
#
par( pty="s" )
plot( out$FP, out$TP, xlab = "1-Especificidade", ylab = "Sensibilidad",
      type="n")
#
##### CI first
#
shade.space <- T
if( shade.space ){
  shadeCI <- function( x, upper, lower ){
    newX <- c(x, x[order(-x)])
    newY <- c(upper, lower[order(-x)])
    polygon(newX, newY, col = "grey", border = NA)
  }
  shadeCI(FP.save, upper, lower)
}

par(new = TRUE)

```

```
risksetROC(b.id$ttot, status = b.id$exit, marker = coxfe0$linear.predictors,  
           predict.time = 1, col = 1, xlab = " ", ylab = "")  
abline(0, 1, lty=2)  
title( "I/D ROC a 1 ano")
```

# Bibliografía

- [1] Cox DR (1972). *Regression models and life-tables*. Journal of the Royal Statistical Society. Series B (Methodological) 34: 187-220.
- [2] Harrell FE Jr (2016). *rms: Regression Modeling Strategies*. R package version 4.5-0. Disponible en: <https://cran.r-project.org/package=rms>. Accedido por última vez o 2 de setembro de 2016.
- [3] Heagerty PJ, Lumley T, Pepe MS (2000). *Time-Dependent ROC Curves for Censored Survival Data and a Diagnostic Marker*. Biometrics 56: 337-344.
- [4] Heagerty PJ, Zheng Y (2005). *Survival Model Predictive Accuracy and ROC Curves*. Biometrics 61: 92-105.
- [5] Heagerty PJ, Saha-Chaudhuri P (2012). *RisksetRoc: Riskset ROC curve estimation from censored survival data*. R package version 1.0.4. Disponible en: <https://cran.r-project.org/package=risksetROC>. Accedido por última vez o 2 de setembro de 2016.
- [6] Pencina MJ, D'Agostino RB Sr, Steyerberg EW (2011). *Extensions of net reclassification improvement calculations to measure usefulness of new biomarkers*. Statistics in Medicine 30: 11-21 .
- [7] Rizopoulos D (2012). *Joint Models for Longitudinal and Time-to-Event Data*. CRC Press.
- [8] Steyerberg EW (2009). *Clinical Prediction Models: A Practical Approach to Development, Validation and Updating*. Springer-Verlag, New York.
- [9] Steyerberg EW, Vickers AJ, Cook, NR, Gerds T, Gonen M, Obuchowski N, Pencina MJ, Kattan MW (2010). *Assessing the Performance of Prediction Models: A Framework for Traditional and Novel Measures*. Epidemiology 21: 128-138 .
- [10] Therneau TM, Grambsch PM (2000). *Modeling Survival Data: Extending the Cox Model*. Springer-Verlag, New York.
- [11] Therneau TM, Lumley T (2009). *Survival: Survival Analysis*. R package version 2.39-5. Disponible en: <https://cran.r-project.org/package=survival>. Accedido por última vez o 2 de setembro de 2016.
- [12] Vickers AJ, Elkin EB (2006). *Decision curve analysis: a novel method for evaluating prediction*. Medical Decision Making 26: 565-574 .