



DURACIÓN DE LA INCAPACIDAD TEMPORAL ASOCIADA A DIFERENTES PATOLOGÍAS EN TRABAJADORES ESPAÑOLES

Capítulo 1 (parte 1)

RESPONSABLE: D^a Eva Calvo Bonacho

Investigación financiada mediante subvención recibida de acuerdo con lo previsto en la Orden TIN/1902/2009, de 10 de julio (premios para el Fomento de la Investigación de la Protección Social –FIPROS-)

La Seguridad Social no se identifica con el contenido y/o conclusiones de esta investigación, cuya total responsabilidad corresponde a sus autores.

Ibermutuamur

Duración de la Incapacidad Temporal asociadas a diferentes patologías en trabajadores españoles.

Capítulo 1:

Descripción y análisis de la BBDD general del proyecto

Esta investigación ha sido financiada al amparo de lo previsto en la Orden TIN/1902/2009, de 10 de junio, (Premios FIPROS, Fondo para el Fomento de la Investigación de la Protección Social).

Tabla de contenidos

Tabla de contenidos	2
Índice de figuras.....	3
Índice de tablas	8
Índice de diagramas	9
Índice de árboles de decisión.....	9
Equipo de investigadores	10
Financiación.....	10
Resumen	11
Introducción	12
Antecedentes.	12
Finalidad y justificación.	16
Objetivos.....	17
Diseño, sujetos y métodos.....	17
Diseño del estudio.	17
Población objetivo del estudio.....	17
Descripción de la muestra	18
Métodos de recogida de datos	23
Diagnósticos que se incluirán en el estudio (según CIE-9)	26
Variables que se van a estudiar	27
Variable dependiente objetivo del estudio.....	28
Variable ocupación (según CNO94) y clasificación en “trabajador manual/no manual”	28
Cálculo de los estándares de duración de la IT en Ibermutuamur	29

Estrategia, métodos y herramientas para el análisis estadístico.....	30
Datos perdidos, selección de variables más relevantes y dependencias y asociaciones entre las mismas	40
Limitaciones y posibles sesgos del estudio.....	68
Calendario previsto para el estudio	69
Plan de trabajo	69
Cumplimiento de las exigencias éticas y legales	70
Resultados del estudio.....	71
Descripción de las variables y análisis de la relación entre las mismas.....	71
La Incapacidad Temporal por Contingencia Común (ITCC)	102
Comparaciones entre dos diagnósticos	152
Minería de datos y modelo predictivo	171
Discusión	191
Descripción de las variables y de sus relaciones.....	192
Análisis de las duraciones por ITCC.....	194
Plazos y componentes de la ITCC	199
Comparación entre dos diagnósticos	202
Minería de datos, modelo predictivo y ERIT	204
Conclusiones.....	207
Bibliografía	209
Anexos	213

Índice de figuras

Figura 1. Frecuencias relativas por sexo.....	18
Figura 2. Histograma de la edad.	19
Figura 3. Distribución, mediana y cuartiles de la edad.....	19
Figura 4. Pruebas de ajuste de la distribución.	20

Figura 5. Relaciones entre la edad y el sexo.	21
Figura 6. Edad y sexo. Frecuencias absolutas	21
Figura 7. Edad y sexo. Frecuencias relativas	21
Figura 8. Distribuciones de la edad por sexo.	22
Figura 9. Cuaderno electrónico de recogida de datos.	23
Figura 10. Componentes de la ITCC.	33
Figura 11. Fases del ciclo de vida de un proyecto de minería de datos según el método CRISP-DM.	35
Figura 12. Secuencia de tareas para el análisis de los datos.	36
Figura 13. Variables con datos perdidos.	41
Figura 14. Casos con datos perdidos	43
Figura 15. Número de variables con algún caso perdido para cada paciente.	44
Figura 16. Histograma del número de variables con datos perdidos.	44
Figura 17. Celdas con datos perdidos.	45
Figura 18. Patrones de valores perdidos y su frecuencia para las variables más relevantes.	46
Figura 19. Valores perdidos en <i>fecha de capacidad laboral</i> y su relación con el diagnóstico.	50
Figura 20. Valores perdidos en <i>fecha de capacidad laboral</i> y su relación con la provincia.	50
Figura 21. Proporción de valores perdidos en <i>fecha de capacidad laboral</i> por diagnóstico.	51
Figura 22. Proporción de valores perdidos en <i>fecha de capacidad laboral</i> por provincia.	51
Figura 23. Valores perdidos y válidos para <i>fecha de capacidad laboral</i>	54
Figura 24. Comparación de los patrones de valores perdidos y válidos para <i>fecha de capacidad laboral</i>	54
Figura 25. Relación de la duración de las bajas y los valores perdidos en <i>fecha de capacidad laboral</i>	55
Figura 26. Comparación de los patrones de valores perdidos y válidos para <i>sexo</i>	56
Figura 27. Distribuciones de la duración de la ITCC y su relación con los valores perdidos para la variable <i>sexo</i>	57
Figura 28. Comparación de los patrones de valores perdidos y válidos para <i>edad</i>	57
Figura 29. Distribuciones de la duración de la ITCC y su relación con los valores perdidos para la variable <i>edad</i>	58
Figura 30. Comparación entre los patrones de valores perdidos y válidos para <i>provincia</i> y de las distribuciones para las duraciones de la IT para ambos subgrupos.	59
Figura 31. Importancia de los factores por su efecto sobre la duración de las bajas por ITCC.	63
Figura 32. Duraciones medias de las bajas por grupo diagnóstico.	64
Figura 33. Distribuciones de las duraciones de las bajas por grupo diagnóstico.	64
Figura 34. Frecuencias por edad y grupo diagnóstico.	65
Figura 35. Frecuencias de las duraciones de las bajas por grupo diagnóstico y sexo.	66
Figura 36. Distribuciones de las duraciones de las bajas por grupo de actividad económica.	67
Figura 37. Duraciones medias de las bajas por grupo de actividad económica.	68
Figura 38. Plan de trabajo.	70

Figura 39. Frecuencias relativas de los niveles de estudio.....	72
Figura 40. Histograma del nivel de estudios por edad categorizado por sexo.....	73
Figura 41. Histograma del número de hijos.	74
Figura 42. Diagrama del número de hijos por edad.	74
Figura 43. Histograma del estado civil.....	75
Figura 44. Frecuencias relativas del estado civil categorizado por sexo.....	76
Figura 45. Histograma de la nacionalidad.....	76
Figura 46. Histograma de casos por provincia.	77
Figura 47. Histograma de casos por CC.AA.	78
Figura 48. Frecuencias relativas por sector de actividad económica.	78
Figura 49. Número de observaciones por sector de actividad y por edad categorizados por sexo.....	79
Figura 50. Número de observaciones por grupo de actividad económica.	80
Figura 51. Número de observaciones por grupo de actividad y edad categorizadas por sexo.	81
Figura 52. Frecuencias relativas comparadas para hombre y mujer para cada grupo de actividad.....	82
Figura 53. Frecuencias relativas para la ocupación (según CNO93)	83
Figura 54. Frecuencias relativas para trabajadores “manuales” y “no manuales”.	84
Figura 55. Distribución de los casos por ocupación, sexo y edad.....	85
Figura 56. Frecuencias absolutas comparadas para ocupación y nivel de estudios.	85
Figura 57. Distribución de los trabajadores “manuales” y “no manuales” por sexo y edad.....	86
Figura 58. Frecuencias relativas para el tipo de contrato.	87
Figura 59. Frecuencias absolutas por edad, sexo y tipo de contrato.....	87
Figura 60. Distribución de los casos por años de antigüedad en la empresa.....	88
Figura 61. Histograma por años de permanencia en la empresa.	88
Figura 62. Antigüedad en la empresa, por edad y sexo.....	89
Figura 63. Histograma de la base reguladora.	90
Figura 64. Distribución de los valores de la base reguladora por edad y categorizados por el sexo.....	91
Figura 65. Distribución de la base reguladora por ocupación y sexo.	92
Figura 66. Frecuencias absolutas de observaciones de la base reguladora por nivel de estudios y sexo	93
Figura 67. Modos de pago.	94
Figura 68. Medios de transporte hasta el puesto de trabajo.....	95
Figura 69. Distancia hasta el centro de trabajo.....	95
Figura 70. Número de observaciones para cada diagnóstico (según CIE-9).	96
Figura 71. Número de observaciones por diagnóstico, edad y sexo.	97
Figura 72. Comorbilidad.	98
Figura 73. Médico que le dio la baja.....	99
Figura 74. Frecuencias de pacientes derivados.....	100
Figura 75. Distribución del número de bajas previas.	102

Figura 76. Distribuciones por sexo del número de bajas previas.	103
Figura 77. Frecuencias relativas del nº de bajas previas por sexo.	104
Figura 78. Relación del número de bajas con la edad y el sexo.	104
Figura 79. Histograma del nº de días de baja previos.	105
Figura 80. Relación entre los días de baja previos, la edad y el sexo.	106
Figura 81. Distribución de la duración de las bajas por ITCC.	107
Figura 82. Duración de las bajas por ITCC.....	108
Figura 83. Duración de las bajas por ITCC comparando hombres y mujeres.	109
Figura 84. Relación de las duraciones de las bajas con la edad y el sexo.	110
Figura 85. Duración de las bajas por ITCC y su relación con intervalos de edad.	111
Figura 86. Duración de las bajas por ITCC y su relación con intervalos de edad y con el sexo.....	112
Figura 87. Duraciones de las bajas por ITCC para cada diagnóstico (CIE-9) categorizadas por sexo. ...	113
Figura 88. Duraciones de las bajas por ITCC para cada diagnóstico.	114
Figura 89. Medias e IC al 99% de las duraciones de las bajas laborales.	115
Figura 90. Duraciones de las bajas laborales para los 10 diagnósticos seleccionados.	115
Figura 91. Efecto de la comorbilidad sobre la duración de las bajas laborales.	117
Figura 92. Diferencias en las duraciones medias de las bajas laborales con ausencia y presencia de comorbilidad	117
Figura 93. Frecuencias de las duraciones de las bajas laborales por nivel de estudios.	118
Figura 94. Distribuciones de las duraciones de las bajas laborales por niveles de estudios.....	119
Figura 95. Duraciones medias de las bajas para cada nivel de estudios.....	119
Figura 96. Duraciones medias de las bajas por provincia.	120
Figura 97. Duraciones de las bajas laborales con relación a los estándares.	122
Figura 98. Relación entre las duraciones de la IT y la causa de alta.	124
Figura 99. Duraciones medias de la IT por causa de alta.	124
Figura 100. Relaciones entre la base reguladora, las duraciones de las bajas laborales y el sexo.	125
Figura 101. Duraciones medias de las bajas por cada uno de los sectores económicos.....	126
Figura 102. Duraciones medias de las bajas por tipo de contrato.....	127
Figura 103. Duraciones de las bajas laborales por modo de pago.	128
Figura 104. Duraciones medias de las bajas laborales por modo de pago.	128
Figura 105. Duraciones medias por tipo de transporte.	129
Figura 106. Lugares donde se administraron los tratamientos recogidos en el protocolo del estudio.	131
Figura 107. Relación entre las duraciones medias de las bajas y las derivaciones a especialistas.	132
Figura 108. Relaciones entre la comorbilidad y el lugar donde se administró el tratamiento.....	132
Figura 109. Estacionalidad de las duraciones de las bajas laborales.....	133
Figura 110. Distribuciones de las duraciones de las bajas con relación a los plazos.	136
Figura 111. Esquema de los componentes de la ITCC.....	137

Figura 112. Proporciones de la suma total las duraciones que corresponden a cada uno de los componentes.	138
Figura 113. Distribuciones de las duraciones para cada uno de los componentes de la ITCC.....	139
Figura 114. Proporciones de los costes totales de las bajas que corresponden a cada uno de los componentes.	140
Figura 115. Costes de los componentes de las bajas por ITCC.	140
Figura 116. Proporciones de los componentes de la baja por ITCC.....	141
Figura 117. Diagrama ternario que representa las proporciones correspondientes a cada uno de los componentes de las bajas por ITCC.	142
Figura 118. Duraciones medias de la lista de espera y de la prolongación por sexo.....	145
Figura 119. Diagramas ternarios para los grupos que superan y no superan los estándares de duración de Ibermutuamur.....	146
Figura 120. Duraciones medias de los componentes de la baja para los pacientes que superan y no superan los estándares de Ibermutuamur.	147
Figura 121. Diagramas ternarios para los deciles de la duración de las bajas por ITCC.....	148
Figura 122. Distribuciones de las duraciones de los componentes de las bajas categorizadas por plazos.	149
Figura 123. Duraciones medias para los plazos de la ITCC por <i>Lista de espera y Prolongación</i>	149
Figura 124. Relación entre las duraciones medias de las bajas y el <i>lugar donde se administró el tratamiento</i>	151
Figura 125. Dispersión de los valores de la duración de las bajas por ITCC para ambos diagnósticos.	153
Figura 126. Comparación de las distribuciones de las duraciones de la IT.	154
Figura 127. Distribuciones de las duraciones de las bajas por ITCC para cada diagnóstico.	154
Figura 128. Frecuencias de ambas patologías en mujeres y hombres.....	155
Figura 129. Efectos de las variables <i>sexo</i> y <i>diagnóstico</i> sobre las duraciones de las bajas por ITCC.	156
Figura 130. Histogramas comparados de las edades para cada una de los diagnósticos.	157
Figura 131. Frecuencias para cada par de valores de edad y duración de la IT categorizadas por sexo para cada uno de los diagnósticos.	158
Figura 132. Frecuencias para cada par de valores de edad y duración de la IT categorizadas por diagnóstico y sexo.....	158
Figura 133. Medias e IC (99%) de las duraciones de la IT categorizadas por diagnósticos, intervalos de edad y sexos.	160
Figura 134. Componentes de la duración de las bajas por ITCC para cada diagnóstico.	161
Figura 135. Distribuciones de los casos que superan y no superan el estándar de Ibermutuamur para cada patología.	162
Figura 136. Medias comparadas de las duraciones de las bajas por cada diagnóstico con relación al estándar.	163
Figura 137. Relaciones entre el diagnóstico y el nivel de estudios.....	164

Figura 138. Relaciones entre los diagnósticos y el estado civil de los pacientes.....	165
Figura 139. Asociaciones entre el tipo de contrato y diagnóstico.	166
Figura 140. Relaciones entre ambos diagnósticos y la ocupación.....	167
Figura 141. Relaciones entre ambos diagnósticos y el sector de actividad económica.....	168
Figura 142. Frecuencias absolutas para trabajadores “Manuales” y “No manuales” categorizadas por el diagnóstico.	169
Figura 143. Asociación entre los diagnósticos y la provincia.	170
Figura 144. Frecuencias de casos por niveles de riesgo.....	181
Figura 145. Frecuencias de casos para cada grupo de riesgo.....	182
Figura 146. Distribuciones de las duraciones para cada grupo de riesgo.	183
Figura 147. Duraciones medias de las bajas para cada grupo de riesgo.....	184
Figura 148. Frecuencias por niveles de riesgo para cada diagnóstico.	185
Figura 149. Costes de las bajas por nivel de riesgo.	186
Figura 150. Costes medios de las bajas por nivel de riesgo.	187

Índice de tablas

Tabla 1. Ocupaciones según CON 94	29
Tabla 2: Tabla de datos perdidos	42
Tabla 3. Valores que modifican la frecuencia de datos perdidos en fecha de capacidad laboral	49
Tabla 4: prueba T con estimaciones separadas de la varianza.....	53
Tabla 5. Relación entre los valores perdidos de <i>sexo</i> y <i>edad</i>	58
Tabla 6. Comorbilidad.	98
Tabla 7. Médico que le dio la baja.....	99
Tabla 8. Frecuencias de pacientes derivados.....	100
Tabla 9. Días de baja por ITCC en los 5 años previos.....	105
Tabla 10. Medias y DE de las duraciones por provincia.....	121
Tabla 11. Análisis descriptivo de las duraciones de las bajas laborales por ITCC.....	138
Tabla 12. Análisis descriptivo de los costes de las bajas laborales por ITCC.....	140
Tabla 13. Aciertos y errores en las predicciones.	179
Tabla 14. Escala de puntuaciones del Riesgo IT (ERIT).	180
Tabla 15. Frecuencias de casos para cada grupo de riesgo.....	181
Tabla 16. Estadísticos descriptivos para cada grupo de riesgo.	184
Tabla 17. Estadísticos descriptivos de los costes de las bajas.....	187
Tabla 18. Tabla de reglas predictivas para cada nivel del ERIT.	190

Índice de diagramas

Diagrama 1. Representación en un diagrama ternario de los componentes de la ITCC.....	143
Diagrama 2. Relación de los componentes de la ITCC con el sexo.	144

Índice de árboles de decisión

Árbol 1. Modelo con un árbol de clasificación tipo CHAID y la variable dependiente continua <i>Duración de la IT</i>	175
Árbol 2. Modelo con un árbol de clasificación tipo CHAID y la variable dependiente dicotómica <i>Superan estándar de Ibermutuamur</i>	176

Anexos

Anexo 1. Las 100 patologías seleccionadas para el estudio.....	213
Anexo 2. Las 10 patologías seleccionadas para su análisis.	216
Anexo 3. Libro de variables.	218
Anexo 4. Matriz de los valores p de las pruebas Chi cuadrado entre variables categóricas.....	223
Anexo 5. Matriz de valores p de los análisis de la varianza entre las variables continuas y categóricas.	224
Anexo 6. Tabla de correlaciones entre variables continuas.	225
Anexo 7. Matriz de covarianzas.	226
Anexo 8. Tabla con los diagnósticos (CIE-9) agrupados.....	227
Anexo 9. Actividades económicas de la empresa (CNAE) agrupadas.	231
Anexo 10. Estándares de duración de Ibermutuamur para cada categoría del CIE-9.....	247
Anexo 11. Prevalencias de los diagnósticos (según CIE-9) encontrados en la muestra.....	248
Anexo 12. Distribución de la base reguladora por sexo.	251
Anexo 13. Ejemplo de código en lenguaje SQL para el modelo predictivo.....	253

Equipo de investigadores

El equipo científico premiado con el FIPROS 2009, que se ha responsabilizado de la realización del presente estudio, estuvo integrado por los siguientes investigadores:

Dra. Eva Calvo Bonacho, Directora Ejecutiva del Proyecto¹
Dr. Arturo González Quintela, Miembro del Comité Científico²
Dr. Miguel Ángel Sánchez Chaparro, Miembro del Comité Científico³
Dr. Carlos Fernández-Labandera, Coordinador del Proyecto¹
D. Luis Quevedo Aguado, Responsable de Documentación¹
Dra. Martha Cabrera Sierra, Colaboradora Científica
Dr. Juan Carlos Sainz Gutiérrez, Responsable de Gestión Datos¹
D. Carlos Catalina Romero, Asesor en Psicología Laboral¹
Dra. Ana Fernández Meseguer, Asesora en Medicina Familiar¹
Dña. Margarita Pérez Castillo, Gestor de datos¹
DUE Montserrat Ruiz Moragas, Asesora en Enfermería Laboral¹

Dichos investigadores están laboralmente adscritos a las siguientes Entidades:

1. Ibermutuamur. Mutua de accidentes de trabajo y enfermedades profesionales de la Seguridad Social nº 274.
2. Universidad de Santiago de Compostela.
3. Universidad de Málaga.

Financiación

La investigación, cuyos resultados se presentan en el presente estudio, ha sido financiada al amparo de lo previsto en la Orden TIN/1902/2009, de 10 de junio, (Premios FIPROS, Fondo para el Fomento de la Investigación de la Protección Social). Premio Fipros 2009/5.

Resumen

En España se vienen realizando esfuerzos para encontrar nuevos modelos de gestión de Incapacidad Temporal por Contingencias Comunes (ITCC) que permitan optimizar el proceso y reducir el gran impacto económico que supone esta prestación social. Los objetivos de esta investigación fueron: establecer estándares de duración de las patologías más frecuentes, compararlos con los establecidos por el Instituto Nacional de la Seguridad Social (INSS) y valorar su posible traslado a los Servicios Públicos de Salud (SPS); analizar los factores (demográficos, laborales, médicos, etc.) que puedan influir en la duración de los procesos de baja; buscar un modelo predictivo y desarrollar herramientas que permitan optimizar la gestión de la ITCC. Se estudiaron 16.174 casos que recibieron la baja por CC entre los años 2003 y 2007 utilizando técnicas de Análisis Exploratorio de Datos y de Minería de Datos. Hemos calculado unos valores estándar de la duración de la ITCC para las patologías más prevalentes basándonos en nuestra experiencia histórica con la ITCC. De los análisis de los factores se desprende el carácter multifactorial y dinámico de los efectos de los mismos sobre la ITCC, así como la importancia dispar que tiene cada uno de ellos. Hemos comprobado que es factible crear modelos predictivos fiables que permitan optimizar la gestión de la ITCC y la necesidad de ajustar y actualizar periódicamente el mismo adaptándolo a los cambios. También hemos analizado en detalle los 10 diagnósticos más relevantes en nuestro estudio. Cabe destacar de esta investigación los estándares de duración y los métodos de análisis propuestos así como el desarrollo de modelos predictivos y de herramientas para optimizar la gestión de la ITCC.

Introducción

Antecedentes.

La Incapacidad Temporal por Contingencias Comunes (ITCC) se puede definir según el artículo 128 del Texto Refundido de la Ley General de la Seguridad Social (TRLGSS) como “la situación en la que se encuentra el trabajador que, por causa de enfermedad común o accidente no laboral, esté impedido para el trabajo y necesite la asistencia sanitaria de la Seguridad Social”. Durante el periodo de tiempo que dure la ITCC el trabajador precisa percibir una prestación económica que sustituya a su salario y una atención sanitaria que reduzca, en lo posible, la duración de su proceso de baja.

Las contingencias que causan ITCC generan un coste social y económico elevado (1; 2). El Instituto Nacional de la Seguridad Social (INSS) destinó para este concepto un total de 8.144 millones de euros (3) en el año 2009. Los Presupuestos Generales del Estado para 2011 incluyen una partida de más de 11.580 millones de euros para los subsidios de IT y otras prestaciones económicas de la Seguridad Social. A estos gastos hay que añadir los costes de la prestación sanitaria, la pérdida de productividad y los costes humanos y familiares provocados por la incapacidad del trabajador (2; 4).

El médico de atención primaria es quien prescribe la baja laboral por enfermedad común en España y también quien decide el momento más oportuno para su reincorporación al puesto de trabajo (5; 6). No existen criterios definidos para determinar la duración de la incapacidad temporal por contingencia común (ITCC). El médico debe decidir esta duración según su experiencia personal (7) que, a su vez, estará influida por diversos factores como la situación laboral del paciente, su propio perfil profesional (8), su antigüedad en la plaza (9) y otros factores.

Anteriormente se ha propuesto la creación de estándares de duración media de los principales procesos causantes de ITCC (10). Algunos autores han determinado valores de referencia y sugieren la realización de otros estudios que continúen esta labor (14).

Hemos encontrado en la bibliografía revisada más relevante 7 estudios publicados (12; 13; 14; 15; 16; 17; 18) que intentan establecer estos estándares, si bien todos ellos evalúan una población reducida tanto en número como geográficamente. Se trata, además, de estudios retrospectivos que no tienen en cuenta diversos factores que pudieran influir en la duración de las ITCC, como pudieran ser los de tipo familiar, económicos, sociales, culturales, laborales, etc.

No conocemos ningún estudio publicado en la actualidad en el que se hayan establecido los tiempos estándar de duración basándose en el seguimiento prospectivo e individual de cada paciente (en baja por ITCC) y en la estimación del momento óptimo para que dicho paciente retorne a su actividad laboral. Solamente encontramos publicaciones basadas en estudios retrospectivos o por consenso.

Para cubrir este vacío, la Dirección General del Insalud ha publicado una guía que pretende servir de orientación al médico del Sistema Público de salud (SPS). También el Instituto Nacional de la Seguridad Social ha publicado recientemente una guía de estándares de duración (19). Ambas guías fueron realizadas por consenso de un grupo de expertos de dicha materia.

Otro punto que consideramos importante evaluar es la fiabilidad diagnóstica emitida en los partes de baja por el SPS. Entre la bibliografía que hemos revisado, encontramos 4 estudios publicados que tratan de analizar el nivel de concordancia diagnóstica y la fiabilidad de los diagnósticos emitidos por el SPS y los factores que puedan estar influyendo (8; 20; 21; 22).

Para asegurar la fiabilidad diagnóstica, los médicos de Ibermutuamur realizaron valoraciones clínicas de los trabajadores durante todo el proceso de seguimiento de la ITCC. Así solicitaron pruebas complementarias, interconsultas a especialistas y tramitaron tratamientos para llegar a un diagnóstico adecuado y a una más pronta recuperación del trabajador.

Factores que influyen en la duración de los procesos de ITCC.

En la duración de los procesos de IT influyen además del diagnóstico y la severidad del cuadro clínico, otra serie de factores (23; 24; 25). En primer lugar, factores sociodemográficos como el sexo, la edad, el estado civil y el nivel de estudios

(13; 14; 15; 24; 26; 27; 28). En segundo lugar, factores sociolaborales como el régimen de Seguridad Social, el grado de satisfacción en el trabajo, el horario y el tipo de trabajo o de contrato (29; 30; 31; 32). Por último, factores sociosanitarios, como las características del médico (8; 33), el tipo de modelo de atención primaria (20; 34) y los problemas de acceso a la atención especializada (35). La dificultad para determinar con certeza la capacidad laboral ante determinadas patologías, hace que los factores citados sean en muchas ocasiones más influyentes en la decisión de volver al trabajo que la propia enfermedad (23). No obstante, en estudios con población española son raras las excepciones en las que se ha tratado de ajustar el efecto de potenciales factores de confusión al evaluar los factores que influyen en la duración de la IT (9).

Los procesos de IT que duran más de un mes consumen gran parte de la prestación económica dedicada a la cobertura de esta contingencia, a pesar de que por número de procesos representan solo un pequeño porcentaje (entre el 10 y el 40%) del total (2; 5; 9; 13; 36; 37). Los resultados de un estudio, que se diseñó para evaluar las IT de larga duración, mostraron que solo el 16% de los procesos de IT acumulaban el 49% de los días de trabajo perdidos (38). Se observó, además, que, de forma sistemática, los diagnósticos de los certificados de IT no justificaban unas duraciones tan largas de incapacidad para trabajar como las detectadas. Por otra parte, en los últimos años se está produciendo en España un aumento progresivo e injustificado de la duración media de la IT, pasando de cifras que rondaban los 35 días (7; 9; 26; 39) durante la década de los 80 hasta unas cifras que se sitúan alrededor de los 45 días durante la década de los 90 (20; 40; 41; 42).

Papel de las Mutuas (MATEPSS)

Las Mutuas de accidentes de trabajo y enfermedades profesionales contribuyen a paliar los efectos negativos que la situación de IT genera y que pueden afectar tanto a nuestro sistema de la Seguridad Social como a otras partes interesadas (empresa y trabajadores). Por una parte, las mutuas contribuyen a disminuir el gasto que constituye la incapacidad temporal en nuestra partida presupuestaria, y por otro, garantizan la incorporación inmediata de aquellos trabajadores con bajas simuladas o cuya duración puede aminorarse con un mayor control médico por parte de los facultativos de estas entidades.

Para poder acceder al subsidio de IT el trabajador debe: estar dado de alta en la Seguridad Social, causar baja laboral por enfermedad (común o accidente no laboral en el caso de la ITCC) recibiendo asistencia sanitaria, y tener un mínimo tiempo de cotización; salvo que se trate de un accidente. Es fundamental que el trabajador no pueda desempeñar su trabajo habitual a causa de su enfermedad.

Desde 1995 las Mutuas, como entidades colaboradoras de la Seguridad Social (SS), tienen las siguientes funciones en la IT por CC:

1. La gestión de la IT por enfermedad común y accidente no laboral a partir del día 16 de la baja.
2. La prestación sanitaria a partir del día 16 de la baja, con el consentimiento del trabajador de acuerdo con el artículo 6 del R.D. 575/1997.
3. La revisión del estado de salud del trabajador mientras dure la baja. El trabajador no se podrá negar a realizar las pruebas complementarias, y su negativa infundada puede dar lugar a una propuesta de alta.
4. La posibilidad de hacer propuesta de alta médica a la Inspección de la Seguridad Social.
5. La posibilidad de anular y suspender el pago del subsidio de baja por IT si no se cumple el período de carencia (180 días por enfermedad común, no necesaria para accidente no laboral) de alta a la Seguridad Social o no estar al corriente de cotización.

En este tipo de gestión, las mutuas tienen los mismos derechos que la Seguridad Social y las entidades gestoras, a excepción de la posibilidad de dar el alta médica.

Según la Asociación de Mutuas de Accidentes de Trabajo (AMAT) el número de empresas asociadas a las Mutuas de Accidentes de Trabajo y Enfermedades Profesionales de la Seguridad Social en diciembre de 2008 es de 1.559.577. Los trabajadores afiliados al sector de mutuas ascienden a 14.176.372; lo que supone un 71,4% del total de trabajadores del Sistema Nacional de Salud (SNS). Las Mutuas disfrutan de una posición idónea para facilitar una adecuada reincorporación al trabajo después de un episodio de IT por CC mediante intervenciones sanitarias.

Datos epidemiológicos

Uno de cada cuatro trabajadores percibe esta prestación de Seguridad Social a lo largo de un año. En España, se producen casi cinco millones de bajas laborales por enfermedad o accidentes cada año, consumiendo alrededor de 300 millones de jornadas de trabajo por este concepto.

Finalidad y justificación.

Recientemente en nuestro país se han emprendido distintas acciones para establecer nuevos modelos de gestión de IT. Podemos destacar el trabajo realizado por el Grupo de Investigación sobre El control de la incapacidad temporal formado por Sepere et al. (43) y financiado con una subvención FIPROS 2005 (Ministerio de Trabajo y Asuntos Sociales. Orden TAS/1051/2005, de 12 de abril). En el mismo se aportaron propuestas de reforma del control de la IT que pueden resultar imprescindibles para optimizar el proceso y reducir el gran impacto económico que supone esta prestación social.

Hemos considerado prioritario continuar esta línea de trabajo desde la perspectiva de las Mutuas de Accidentes de Trabajo y Enfermedades Profesionales colaboradoras de la Seguridad Social (MATEPSS) piedras angulares del SNS.

Ibermutuamur como mutua de accidentes de trabajo y enfermedad de la Seguridad Social, nº274, desde el año 1996 viene realizando la prestación económica de los procesos en ITCC de los trabajadores en baja de las empresas que tienen concertada esta contingencia con ella. Para este fin, cuenta con un servicio médico especializado en el control de todos los procesos en baja y en el seguimiento de los pacientes hasta que se produce el alta, que está presente en toda la geografía española y que cuenta con información detallada de los aspectos clínicos, sociales, laborales, familiares y culturales de los procesos y de los pacientes implicados en los mismos.

La implantación territorial y los medios con que cuenta Ibermutuamur nos ofrecían una oportunidad única para estudiar la ITCC y los factores involucrados en la misma en varios de sus aspectos.

Objetivos.

Primario:

Establecer los estándares de duración de la ITCC para las patologías más frecuentes en la población laboral protegida por Ibermutuamur.

Para ello se analizarán los datos recogidos por Ibermutuamur entre los años 2003 y 2007 (inclusivos) correspondientes a los seguimientos individualizados de los procesos realizados por los médicos de Ibermutuamur responsables del control de la incapacidad temporal por Contingencia Común.

Secundarios:

1. Comparar los estándares de duración obtenidos con los establecidos por el INSALUD dentro del “Manual de gestión de la Incapacidad Temporal” (2001) y los propuestos por el INSS en su libro “Tiempos estándar de Incapacidad Temporal” (2009).
2. Valorar la forma de trasladar los estándares de duración de Ibermutuamur a los Servicios Públicos de Salud (SPS).
3. Analizar qué factores pueden incluir en la duración de la ITCC (variables demográficas, socioeconómicas, etc.).

Diseño, sujetos y métodos

Diseño del estudio.

Estudio epidemiológico multicéntrico, observacional descriptivo y prospectivo.

Población objetivo del estudio.

La **población objetivo** de estudio se correspondió con la población activa protegida por Ibermutuamur en lo que se refiere a las Contingencias Comunes (CC). Dicha población, en el periodo en el que se realizó el estudio, incluía alrededor de 632.000 trabajadores.

La **población experimental** la formaron los pacientes (protegidos por Ibermutuamur) que recibieron la baja por Contingencia Común entre los años 2003 y 2007.

Descripción de la muestra

Tamaño muestral

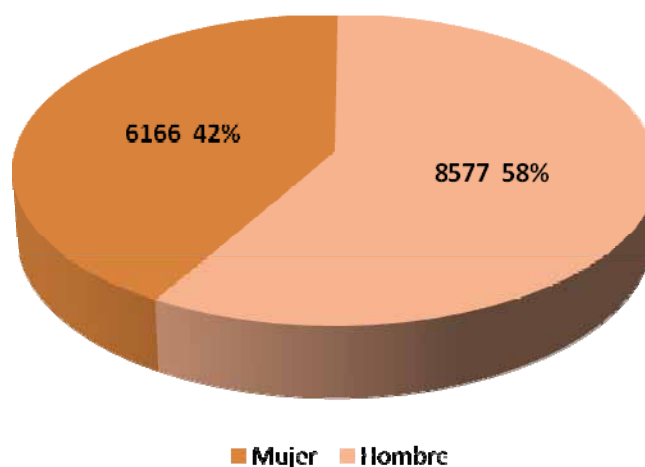
En total se recogió una muestra de **16.174** casos de pacientes (protegidos por Ibermutuamur) que recibieron la baja por Contingencia Común entre los años 2003 y 2007 y que estaban diagnosticados de alguna de las 100 patologías previamente seleccionadas (*Ver "Diagnósticos que se incluirán en el estudio (según CIE-9) en la página 26*). Dichos diagnósticos se seleccionaron entre aquellos que presentaban una incidencia mayor de 150 casos al año, con una duración media mayor de 30 días y representativos de diferentes grupos de patologías.

Edad y sexo de los pacientes incluidos en el estudio

1. Sexo.

La muestra está formada por 6.166 mujeres (42%) y 8.577 hombres (58%) que cumplieron los criterios de inclusión en el estudio descritos anteriormente (Figura 1).

Figura 1. Frecuencias relativas por sexo.



2. Edad

Los pacientes de nuestra muestra tenían una edad media de 38 años (IC 95% de 37,75 a 38,12) con una DE de 11,57.

Por rangos de edad observamos en las prevalencias de la ITCC aumentan de forma progresiva y lineal hasta los 30 años. Éstas alcanzan las frecuencias más elevadas en los tramos entre los 25 y los 40 años, es decir en pacientes más bien jóvenes. La frecuencia va disminuyendo progresivamente y de forma casi lineal a partir de los 35 años hasta la edad de la jubilación (Figura 2 y Figura 3).

Figura 2. Histograma de la edad.

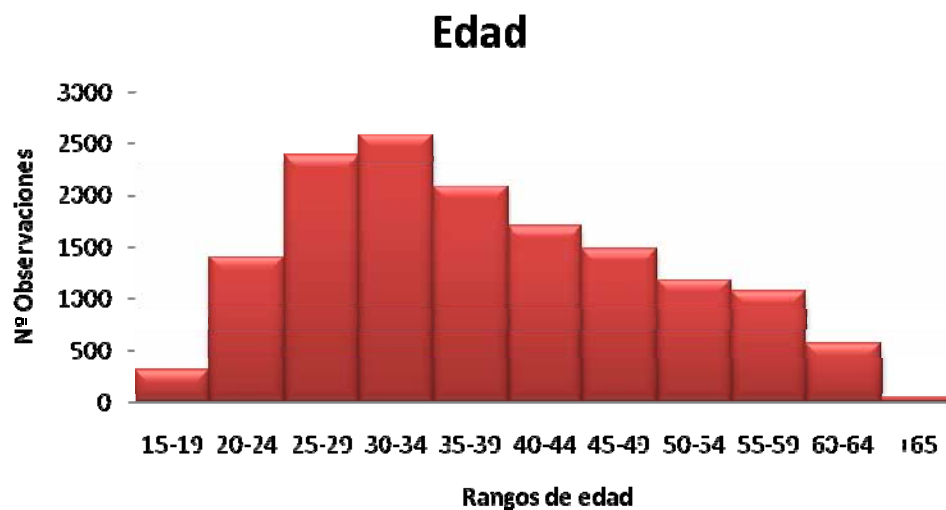
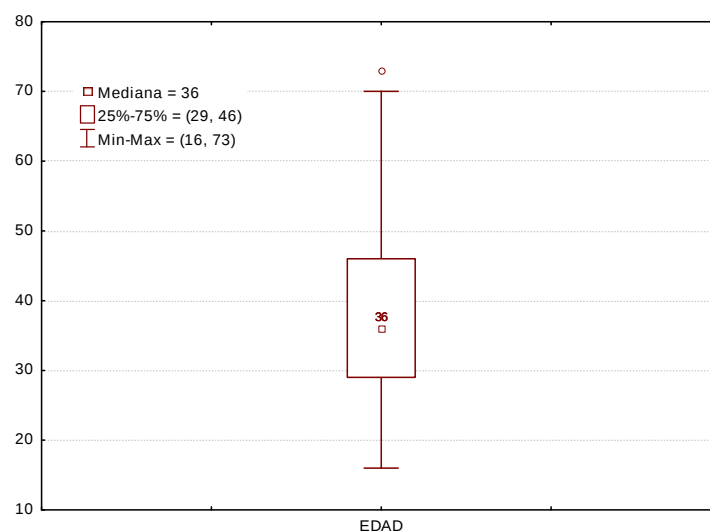
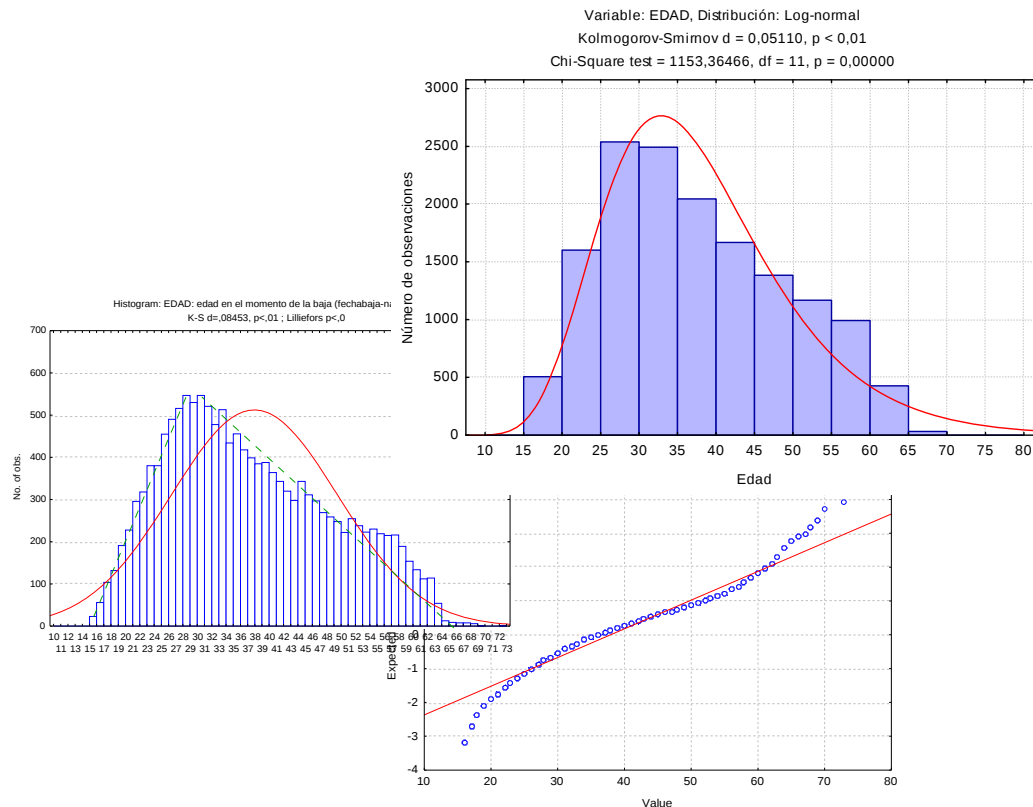


Figura 3. Distribución, mediana y cuartiles de la edad.



Tipo de distribución para la edad: la distribución de la edad en la muestra no cumple los criterios de normalidad (K-S $d=,084$; $p<0,01$) (Figura 4). Se aproxima mucho, aunque tampoco cumple los criterios (K-S $d=0,051$; $p<0,01$), a una distribución log-normal.

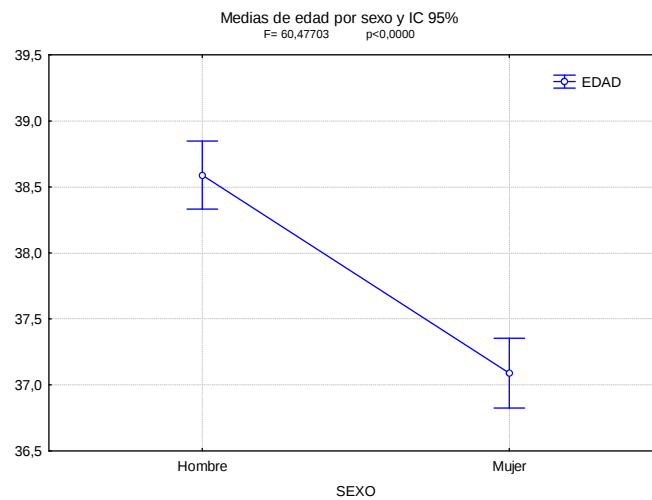
Figura 4. Pruebas de ajuste de la distribución.



Edad por sexo

La edad media para las mujeres de la muestra fue de 37 años (DE 11 años) y para los hombres de 39 años (DE 12 años). Las diferencias entre ambos grupos (Figura 5), en cuanto a la edad, fueron estadísticamente significativas ($p<0,01$).

Figura 5. Relaciones entre la edad y el sexo.



La edad sigue distribuciones bastante parecidas para cada uno de los sexos (Figura 6). Si observamos la distribución de las frecuencias relativas para los diferentes rangos de edad calculadas sobre la frecuencia marginal para cada categoría de sexo (es decir la proporción correspondiente dentro de su propio sexo), observamos que las mujeres presentan mayores frecuencias relativas en los tramos entre 25 y 40 años (muy especialmente entre los 25 y 35 años) y menores en el resto, especialmente en los tramos de mayor edad (Figura 7). Las diferencias por rangos de edad y sexo son en conjunto estadísticamente significativas ($\chi^2=328,54$; $p<0,000$).

Figura 6. Edad y sexo. Frecuencias absolutas

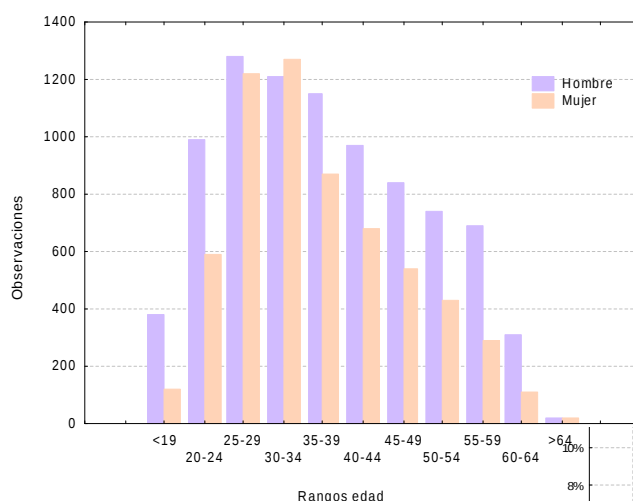


Figura 7. Edad y sexo. Frecuencias relativas

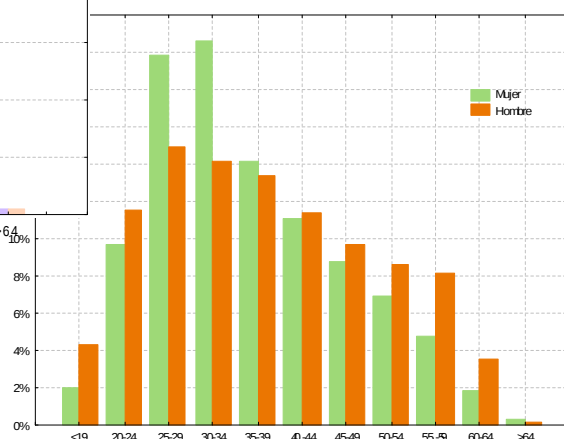
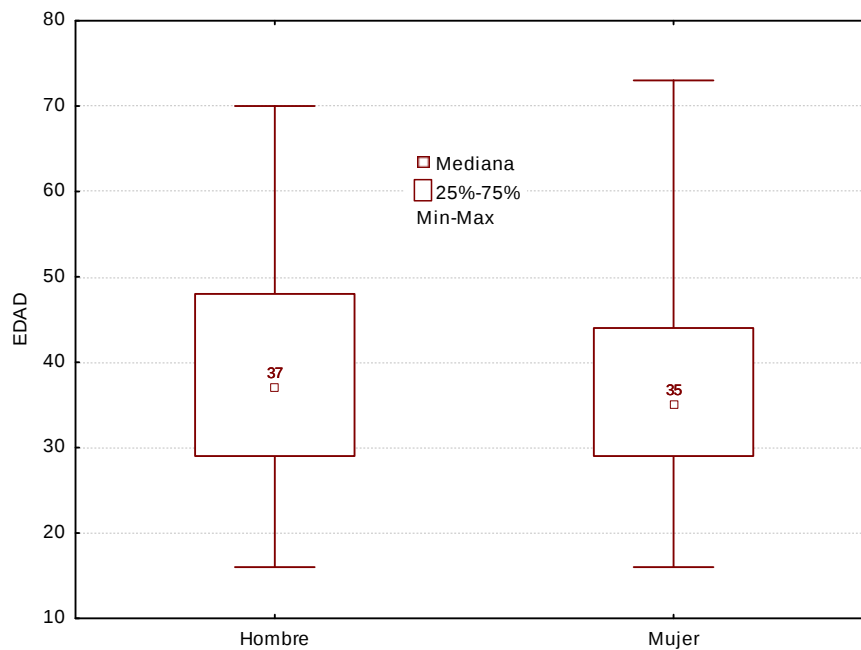


Figura 8. Distribuciones de la edad por sexo.



Criterios de inclusión

1. Pacientes que causaron baja por Incapacidad Temporal (IT) debida a Contingencia Común en el periodo estudiado (enero de 2003 a marzo de 2007)
2. Pacientes diagnosticados de alguna de las patologías seleccionadas que se citan más adelante (siguiendo la clasificación de enfermedades CIE-9). Ver "Diagnósticos que se incluirán en el estudio (según CIE-9) en la página 26.
3. Y que hallan sido valorados presencialmente por un médico de Ibermutuamur.

Método de muestreo

Se procedió a realizar un *muestreo consecutivo* entre la población activa protegida por Ibermutuamur en lo que se refiere a las Contingencias Comunes (alrededor de 632.000 trabajadores). Los 140 investigadores de campo que participaron en el estudio incluyeron en el mismo los primeros 20 casos que acudían cada mes para iniciar el control de la incapacidad temporal y que presentaban alguno

de los diagnósticos objeto del estudio. La recogida de datos fue realizada por médicos de control de la IT en Ibermutuamur entre enero de 2003 y marzo de 2007.

Métodos de recogida de datos

Entrada y gestión informática de los datos

Ciento cuarenta investigadores (médicos de Contingencias Comunes) realizaron un seguimiento presencial individualizado de los pacientes en baja por contingencia común durante el periodo comprendido entre enero de 2003 y marzo de 2007. Utilizaron cuadernos de recogida de datos (CRD) electrónicos para recolectar los datos específicos del estudio (como, por ejemplo, sexo, edad, tipo de contrato, tipo de pago, ocupación, sector de actividad, base reguladora, nivel de estudios, lugar de residencia, organismo que realiza la intervención y tipo de intervención, etc.) y también se recolectaron datos procedentes de la historia clínica electrónica de Ibermutuamur (correspondientes a dichos pacientes). En cada consulta, el médico de control valoró la fecha de capacidad laboral o la fecha de alta óptima en función de los síntomas, signos y ocupación del trabajador. El intervalo máximo entre consultas se estableció en 30 días.

Para facilitar la recogida de los datos se creó un CRD electrónico que se integró con la aplicación informática de Ibermutuamur dedicada a recoger las historias clínicas de los pacientes (ver Figura 9). Se recogieron los siguientes datos:

Figura 9. Cuaderno electrónico de recogida de datos.

The image shows two overlapping software windows from the 'IberPrisma Sanitario' application. The background window is titled 'IberPrisma Sanitario - Iberfajas: Datos de asistencia' and contains fields for patient identification (HI 1507925, Cpx 1, JOSÉ RAMÓN BLANCO ESCOBAR), date of assistance, and checkboxes for 'Capacidad Laboral', 'Empeoramiento', and 'Posible incapacidad por...'. It also has a section for 'Tratamientos' with five numbered buttons. The foreground window is titled 'IberPrisma Sanitario - Iberfajas: Datos de pacientes' and is for 'Agrupación de nueva historia: 1507925'. It contains several sections: 'Datos de la Empresa' (C.C.C. de Empresa, C.N.A.E., Modo de Pago), 'Datos del Trabajador' (Ocupación, Tipo contrato, Régimen S.S., Nivel estudios, Nacionalidad, Estado civil, Modo Transporte, Sexo), 'Datos Administrativos' (Médico Baja S.S., Centro Baja S.S., Médico Control, Provincia origen Baja, Fecha Grabación, Fec. inicio control), and 'Datos Clínicos' (Diagnóstico Baja parte S.S., Incapacidades Permanentes previas concedidas, IT ul. 5 años, Menstrual, Grado minusvalía, Año recon. minus, Ingreso empresa, Fecha nacimiento). Both windows have 'Grabar' and 'Salir' buttons at the bottom.

Datos demográficos y relativos a la IT

1. **H^a Iberprisma:** Identificación del paciente.
2. **Fecha de baja:** fecha en la que el médico del Sistema Público de Salud ha expedido la baja médica por ITCC.
3. **Nº episodios previos:** procesos anteriores tanto por ITCC como por Accidente laboral con el número de días totales de los procesos anteriores que fueron registrados en nuestro sistema informático.
4. **I.T en los últimos 5 años:** se recoge, en días, los periodos que el paciente ha permanecido en situación de baja médica por ITCC en los últimos 5 años.
5. **Fecha de nacimiento:** edad
6. **Sexo**
7. **Estado civil:** se recogen las variables soltero, casado, viudo, divorciado, pareja de hecho, Ns/Nc.
8. **Nacionalidad:** se recogen las variables España, Europa, Iberoamérica, Magreb, resto del mundo, Ns/Nc.
9. **Número de hijos:** numérico.
10. **Ocupación:** se recoge mediante el código Nacional de ocupaciones, a tres dígitos.
11. **CCC de la empresa:** identifica a que empresa pertenece el trabajador
12. **CNAE:** código nacional actividades económicas.
13. **Nivel de estudios:** se recogen las variables sin estudios, estudios primarios, estudios secundarios, titulación media, titulación superior, Ns/Nc
14. **Pluriempleo:** hace referencia a si el trabajador trabaja en más de una empresa.
15. **Minusvalías:** si al trabajador se le ha concedido algún grado de minusvalía y el porcentaje del grado reconocido.

Datos Laborales

1. **Tipo de contrato:** hace referencia a si el trabajador tiene un contrato fijo o eventual.
2. **Tipo de pago:** se distingue entre:
3. **Pago delegado:** identifica a los trabajadores que siguen perteneciendo a la empresa y la prestación económica mediante su proceso de ITCC la realiza la empresa.

4. **Pago directo por cuenta ajena:** identifica a los trabajadores que durante su proceso de baja médica han perdido el vínculo laboral con su empresa y la prestación económica durante el proceso de ITCC lo realiza la mutua.
5. **Pago directo cuenta propia:** Identifica a los paciente que trabajan por cuenta propia y que el abono de la prestación económica de ITCC la realiza íntegramente la mutua desde el día 3º de su baja médica.
6. **Antigüedad en el trabajo:** tiempo que lleva trabajando en esa empresa.
7. **Distancia de su residencia habitual al trabajo** en KM.
8. **Modo de transporte** para ir al centro de trabajo: se recogen las variables (coche conductor, coche pasajero, transporte público, otros, Ns/Nc.
9. **Pluriempleo:** hace referencia a si el trabajador trabaja en más de una empresa.
10. **Minusvalías:** si al trabajador se le ha concedido algún grado de minusvalía y el % del grado reconocido.
11. **Base reguladora:** cantidad económica que percibe el trabajador durante su proceso de ITCC.
12. **Incapacidades Permanentes previas concedidas** y relación con la actual baja.
(preguntar directamente al paciente)

Datos Sanitarios y administrativos

1. **Nº de colegiado** del médico que expidió la baja médica.
2. Si el médico que emitió la baja es **sustituto, titular o especialista**.
3. **Localidad origen baja:** zona geográfica origen de la baja médica.
4. **Área sanitaria:** dentro de la zona geográfica, identifica en que centro de salud se emitió el parte médico de baja.
5. **Medico control:** identifica el médico de Ibermutuamur que realiza el control médico.
6. **Intervalo baja- grabación:** demora en integrar a nuestro sistema informático el parte de baja que nos llega a los servicios médicos de Ibermutuamur.
7. **Intervalo grabación inicio:** Tiempo de demora en el inicio del control sanitario.
8. **Tipo de baja:** hace referencia a si es una baja por enfermedad o accidente no laboral.
9. **Diagnóstico** del parte emitido por el médico del sistema público de salud.

Datos asistenciales

1. **Fecha de la visita presencial:** se recoge el número de veces que se valora al paciente en consulta y el intervalo entre estas valoraciones.
2. **Resultado de la exploración:** se recogen las variables no se explora, mejoría, estable, empeoramiento, no valorable.
3. **Lista de espera para intervención quirúrgica:** se recoge en las patologías quirúrgicas si hay demora en el Servicio Público de Salud (SPS) para la intervención.
4. **Fecha de intervención:** momento en el que se realiza la intervención quirúrgica.
5. **Diagnóstico:** en este campo se codificara el diagnóstico, según protocolo de codificación (Anexo 2), emitido por el médico de control tras la valoración del paciente
6. **Co-diagnósticos.** Posible patología que puedan estar influyendo en el proceso de baja pero que no sea la causa de esta.
7. **Pruebas complementaria,** realizadas por Ibermutuamur en cada visita.
8. **Interconsultas a especialistas** a iniciativa del médico de Ibermutuamur.
9. **Tratamiento:** se codificará según protocolo de codificación y se recogerá por separado si dicho tratamiento se ha realizado en el SPS o en Ibermutuamur.
10. **Capacidad laboral:** El médico de control en cada visita presencial determinará según su criterio si el paciente está capacitado para trabajar o no, si está capacitado se recogerá la fecha en la que el médico considera que el paciente debe incorporarse a su puesto de trabajo.
11. **Posible Incapacidad Permanente:** el médico de control determinará si el paciente tiene posibilidad de incorporarse a trabajar o por el contrario es subsidiario de una Incapacidad Permanente y fecha de esta.
12. **Asistencia final:** Se recogerá la fecha real del alta emitido por el SPS y la causa.

Diagnósticos que se incluirán en el estudio (según CIE-9)

Se establecieron 100 diagnósticos diferentes como objetivos de estudio (Anexo 1, página 215). Los diagnósticos se seleccionaron entre aquellos que presentaban la incidencia más alta y, en todo caso, con más de 150 casos al año. Esto se hizo con la intención de que el estudio pudiera completarse en un plazo de tiempo razonable y, al

mismo tiempo, pudiéramos obtener información de los procesos con prevalencia más alta. También se tuvo en cuenta que todos los diagnósticos presentaran una duración media mayor de 30 días. Por último, se cuidó el hecho de que los diagnósticos seleccionados representaran diferentes grupos de patologías diferenciados por características tales como el carácter propio de las mismas o el tipo de tratamiento que requieren (quirúrgico o no). Para ello se escogieron diagnósticos clasificados en las siguientes categorías:

- Diagnósticos de patología quirúrgica.
- Diagnósticos de patología médica.
- Diagnósticos de patología ortopédica.
- Diagnósticos de patología psiquiátrica.
- Diagnósticos de patología ósteo-articular.
- Diagnósticos de patología del embarazo.

Los diagnósticos seleccionados pueden consultarse en la tabla del Anexo 1.

De estos diagnósticos hemos analizado por separado y en profundidad diez, escogidos entre aquellos que presentaron una mayor incidencia y que representaban diferentes grupos de patologías, según la clasificación que vimos anteriormente. Las patologías seleccionadas fueron las que pueden consultarse en la tabla del Anexo 2.

Variables que se van a estudiar

Se recogieron en el estudio datos correspondientes a 145 variables. De ellas seleccionamos, en un primer paso, las 35 que consideramos eran relevantes para alcanzar los objetivos planteados. El conjunto de las variables seleccionadas estaba formado por las siguientes:

- | | |
|------------------------|-----------------------------|
| 1. Sexo | 2. Nivel de estudios |
| 3. Estado civil | 4. Nacionalidad |
| 5. Edad | 6. Nº de hijos |
| 7. Actividad económica | 8. Antigüedad en la empresa |
| 9. Medio de transporte | 10. Km hasta el trabajo |
| 11. Ocupación | 12. Sector de actividad |

- | | |
|--|-------------------------------|
| 13. Manual/No manual | 14. Pluriempleo |
| 15. Tipo de contrato | 16. Nº de bajas |
| 17. Causa del alta | 18. Duración de la IT |
| 19. Día de la semana en que recibe la baja | 20. Mes en que recibe la baja |
| 21. Duración total de ITs 5 años previos | 22. Provincia |
| 23. Tipo de baja | 24. Nº de co-diagnósticos |
| 25. Quién le ha tratado? (IB, SPS o ambos) | 26. Base reguladora |
| 27. Coste de la IT | 28. Tipo de pago |
| 29. Tipo de médico (titular/sustituto) | 30. Tipo de alta |
| 31. Derivación a otro especialista | 32. Gravedad del diagnóstico |
| 33. Tipo de intervención | 34. Diagnóstico |
| 35. IP previas | |

Posteriormente, después de analizar estas variables y las relaciones entre las mismas, se hará una segunda selección, definitiva, de variables más relevantes.

En el Anexo 3 puede consultarse una tabla (Libro de variables.) que resume información sobre dichas variables, a saber: tipo de variable, rango de valores, algunos estadísticos, valores perdidos, etc.

Variable dependiente objetivo del estudio

La variable objetivo de este estudio, la duración de la IT, es una variable calculada a partir de las fechas de alta y de baja de acuerdo con la siguiente fórmula:

$$\text{dur IT} = \text{fecha de alta} - \text{fecha de baja} + 1 \text{ día}$$

Variable ocupación (según CNO94) y clasificación en “trabajador manual/no manual”

Se ha recogido la variable ocupación siguiendo la Clasificación Nacional de Ocupaciones del año 94 según se relaciona en la web del Instituto Nacional de Estadística (INE) (www.ine.es/clasifi/cnoh.htm). Para simplificar se han tomado las 10 categorías en las cuales se agrupan todas las ocupaciones.

En la Tabla 1 puede verse la relación de dichas categorías y su correspondencia con la consideración de “trabajador manual” o “de cuello azul”, según el término anglosajón, y “no manual” o “de cuello blanco”.

Tabla 1. Ocupaciones según CON 94

0	Fuerzas Armadas	No manual
1	Dirección	No manual
2	Técnicos, científicos e intelectuales	No manual
3	Técnicos y profesionales de apoyo	No manual
4	Administrativos	No manual
5	Restauración, personales, protección y comerciales	Manual
6	Agricultura y pesca - cualificados	Manual
7	Manufacturas, construcción y minería - cualificados	Manual
8	Operadores y montadores	Manual
9	No cualificados	Manual

Cálculo de los estándares de duración de la IT en Ibermutuamur

Uno de los objetivos de este proyecto es establecer la duración estándar óptima para diferentes diagnósticos dentro de Ibermutuamur.

Criterios que hemos seguido.

Como criterio general y para reducir el número de diagnósticos hemos empleado las categorías de la clasificación CIE-9 (códigos solo al nivel de 3 dígitos, incluyendo las subcategorías correspondientes). En el documento monográfico anexo que dedicamos a los estándares de duración hemos realizado un análisis más detallado y calculado éstos también para las subcategorías más relevantes.

Criterio 1: el incluir solamente los casos con IT dentro del plazo habitual (sin prórroga ni agotamiento, es decir 365 días como máximo). Si supera dicho plazo es más probable que haya tenido complicaciones, codiagnósticos, casos en que se alarga la IT por otras causas, etc.

Criterio 2: calculamos la duración utilizando la siguiente fórmula: fecha óptima de alta menos fecha de baja (o de intervención si se trata de un proceso quirúrgico) más un día. La fecha óptima de alta sería la más temprana de dos fechas: la de alta oficial o la propuesta por el médico de control de Ibermutuamur. El médico de control indica en qué fecha el paciente ya está capacitado para trabajar y, por lo

tanto, debería recibir el alta y reincorporarse al trabajo. Ésta con frecuencia es anterior a la fecha de alta que finalmente recibe el paciente pero no siempre es así. En ocasiones el paciente se reincorpora al trabajo antes de la fecha señalada por el médico de Ibermutuamur. Por ello tomamos la más temprana de ambas fechas.

Criterio 3: estimamos la mediana de las duraciones así calculadas y la tomamos como estándar óptimo de Ibermutuamur.

Como hemos comentado ya, para simplificar hemos calculado el estándar para cada categoría de los diagnósticos del CIE-9 (códigos con tres dígitos y sus subcategorías). Los valores estándar de duración de la IT para cada categoría diagnóstica según CIE-9 pueden consultarse en el Anexo 10. Estándares de duración de Ibermutuamur para cada categoría del CIE-9.

Dentro de este proyecto, hemos elaborado un documento monográfico dedicado al estudio de los estándares de duración de la IT en Ibermutuamur para los diagnósticos (según CIE-9) más prevalentes. En dicho documento podrá encontrar el lector un análisis más profundo sobre este tema, así como la comparación de dichos estándares con aquellos fijados por el INS (Instituto Nacional de Salud).

Estrategia, métodos y herramientas para el análisis estadístico

Estrategia del análisis estadístico

Se procedió al análisis de los datos siguiendo un plan de análisis estadístico establecido previamente. Los pasos a seguir fueron:

- A. Análisis global de la B.D. de ITCC (todos los diagnósticos).
- B. Análisis global de la duración y del coste de la ITCC (todos los diagnósticos).
- C. Cálculo de los estándares de duración por proceso patológico en Ibermutuamur y comparación de los mismos con los publicados por el INSS.
- D. Análisis comparativo entre dos diagnósticos más representativos.

- E. Análisis específico de 10 diagnósticos previamente seleccionados en base a su prevalencia y relevancia.

En el presente documento se presentarán los resultados correspondientes a los puntos A, B y D. Los puntos C y E se presentarán en sus correspondientes documentos, a saber: un documento independiente para los estándares de duración de la IT del punto C y un documento por cada uno de los 10 diagnósticos estudiados (10 documentos en total dedicados a analizar en profundidad dichos diagnósticos) a los que hace referencia el punto E. En general se presentarán los resultados siguiendo el siguiente esquema:

1. Exploración previa de los datos. Análisis de los datos perdidos. Análisis de las dependencias, correlaciones y asociaciones entre las variables.
2. Selección de variables y reducción del nº de variables, casos y categorías.
3. Conocimiento de los datos. Análisis exploratorio de los datos (AED).
 - 3.1. Descripción de los datos.
 - 3.2. Análisis de las relaciones entre variables.
 - 3.3. Análisis de la ITCC.
 - 3.4. Comparación entre dos diagnósticos relevantes.
4. Búsqueda de patrones ocultos. Creación y evaluación de un modelo predictivo.
5. Aplicación del modelo.

Herramientas empleadas en el análisis estadístico

Para todas las tareas relacionadas con la recogida de datos, creación y manejo de las bases de datos se han utilizado los servidores y software dedicados a estos fines propiedad de la mutua así como la aplicación Acces de Microsoft Office 2007 y de las herramientas de análisis estadístico que se referencia a continuación.

Para los análisis estadísticos, la creación de los gráficos y los procedimientos de minería de datos se han utilizado diversas aplicaciones, a saber: Microsoft Office Excel 2007, SPSS Statistics 17, SAS System 9 for Windows, Statistica 8, Weka 3.6 y RapidMiner 5.

Análisis exploratorio de los datos (AED)

Se ha realizado con el fin de conocer y entender los datos, su calidad, las variables, la estructura de los datos perdidos y su impacto, las relaciones entre las variables, los patrones ocultos en los datos y la relevancia de las diferentes variables. Permite conocer mejor la B.D., entender la ITCC, seleccionar las variables más relevantes, dirigir y controlar la validación y limpieza de los datos, realizar transformaciones de las variables, crear nuevas variables necesarias para los análisis, descubrir relaciones entre variables, correlaciones y patrones ocultos en los datos.

En el presente documento se presentarán, en el capítulo dedicado a los métodos, los resultados de las siguientes pruebas:

- Análisis de los datos perdidos.
- Análisis de las dependencias, correlaciones y asociaciones entre las variables.
- Selección de variables y reducción del nº de variables, casos y categorías.

Descripción de las variables más relevantes.

Para conocer las variables más relevantes para el estudio, como son, por ejemplo, la duración de la IT, las variables demográficas y laborales, las variables clínicas y otras. Analizaremos sus distribuciones de frecuencias y calcularemos los estadísticos que definen mejor las mismas. Sus resultados se presentarán en el correspondiente capítulo, denominado “Resultados del estudio”, tanto en el documento presente como en el resto de los documentos.

Análisis de las relaciones entre variables.

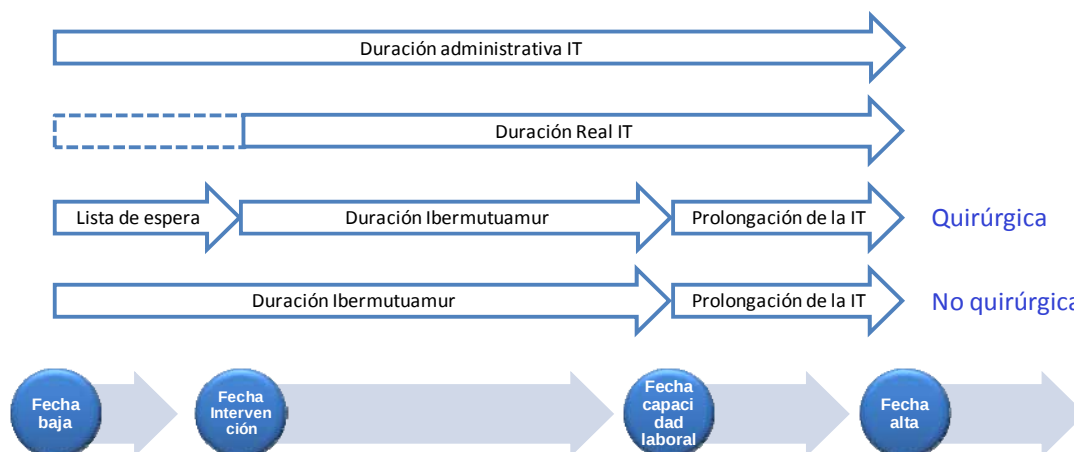
Nos permitirá descubrir qué relaciones existen entre las variables más relevantes y conocer qué influencia pueden tener los diferentes factores sobre la duración de la ITCC. Analizaremos todo tipos de relaciones posibles, como son: asociaciones, correlaciones, covarianzas, relaciones lineales, de clase, etc. Como en el

punto anterior, sus resultados se presentarán en el capítulo denominado “Resultados del estudio”, tanto en el documento presente como en el resto de los documentos.

Análisis de la ITCC y de sus componentes

Se analizará en profundidad la ITCC poniendo especial atención a lo que denominaremos como “componentes de la IT”. Dichos componentes se corresponden con tres tramos de la duración de la ITCC que sumados igualan la duración total de la misma para cada paciente. Hemos definido dichos tramos basándonos en las fechas de baja, de intervención (si se trata de un proceso quirúrgico), de capacidad laboral (estimada por el médico de control de la mutua cuando procede) y de alta. En la Figura 10 presentamos un esquema que muestra los diferentes tramos y cómo se definen éstos.

Figura 10. Componentes de la ITCC.



Los resultados correspondientes a este punto podrán consultarse en el capítulo denominado “Resultados del estudio” del presente documento.

Análisis comparativo entre los dos diagnósticos más representativos

Para conocer algunos aspectos que caracterizan la ITCC resulta muy interesante realizar un análisis comparativo entre dos diagnósticos que presentes

propiedades claramente diferenciadas. Cada proceso patológico presenta unas características diferenciadas con relación a la IT y a diversos factores que influyen en la misma. Así, la duración de la IT será diferente en función de la patología diagnosticada; de su gravedad; del tratamiento que requiera, por ejemplo, quirúrgico o clínico; de que sea más prevalente un sexo o en el otro; o en ciertos rangos de edad; o en ciertas ocupaciones; de que el paciente sea atendido por el médico de Atención Primaria o, bien, derivado a un especialista; etc.

En este estudio, hemos observado que los diagnósticos que se corresponden con los códigos del CIE-9 311 (Trastorno Depresivo no clasificado bajo otros conceptos) y 550 (Hernia Inguinal) representan dos polos claramente diferenciados en base a los perfiles que presentan en relación con la IT y con diversos factores que influyen sobre la misma. Además se han recogido muestras suficientemente grandes para ambos (1.211 casos para el código 311 y 1.015 casos para el 550) ya que se encuentran entre los procesos más prevalentes entre la población trabajadora protegida por Ibermutuamur.

Minería de datos: identificación de patrones en la muestra.

Las modernas técnicas de minería de datos nos ayudarán a buscar y descubrir los patrones de comportamiento ocultos en el conjunto de los datos. Dichos patrones pueden ser explicativos o predictivos de lo que sucede con relación a la frecuencia, duración o del coste de la ITCC.

Método.

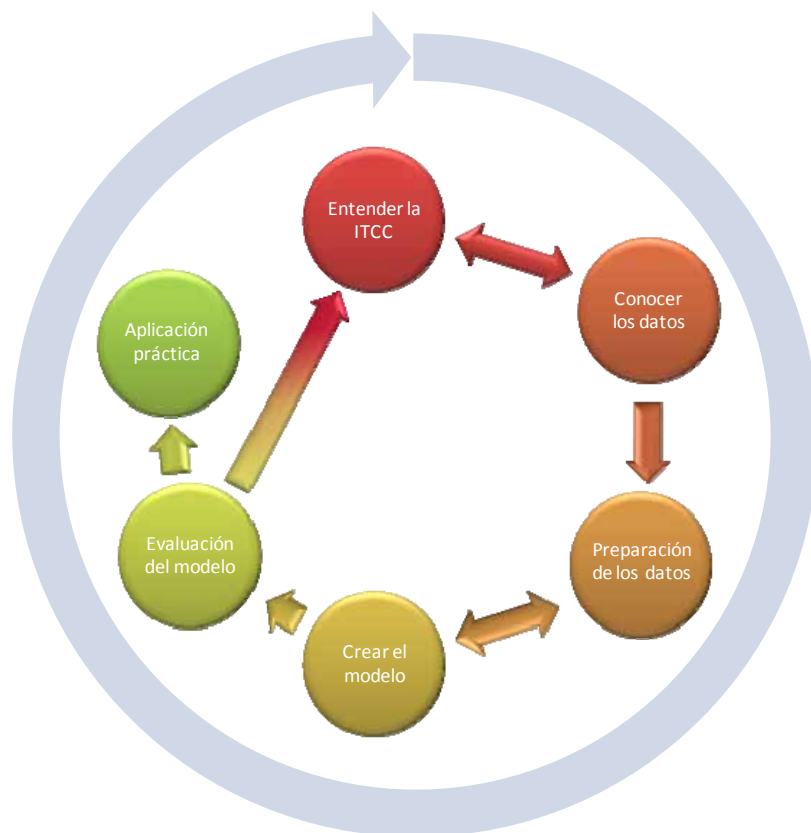
Para planificar y organizar las tareas correspondientes a la minería de datos seguiremos, adaptándolo a nuestros medios y necesidades, el llamado método CRISP-DM.

Este método comenzó a desarrollarse en el año 1996 por algunas compañías que ya tenían alguna experiencia en el empleo de la minería de datos. Dichas compañías fueron: Daimler-Benz (hoy DaimlerChrysler), una de las primeras organizaciones que utilizó la minería de datos en su negocio; ISL (posteriormente conocida como SPSS y que recientemente se ha fusionado con IBM), que lanzó en 1994 la primera aplicación especialmente diseñada para minería de datos

(Clementine); y NCR, que había dedicado equipos de especialistas en minería de datos como valor añadido para sus clientes de grandes almacenes de datos. Todos ellos comenzaron desarrollando la minería de datos utilizando el ensayo y el error. Finalmente, decidieron desarrollar un modelo de procedimiento estándar no propietario y abierto a todo el mundo. Se creó un consorcio y comenzaron los trabajos, con fondos de la Comisión Europea, para desarrollar el que se denominó Cross-Industry Standard Process for Data Mining (CRISP-DM). Se creó el SIG (Special Interest Group) para facilitar la participación de todos aquellos que estuvieran interesados en el proyecto. En el año 2000 se publicaría lo que sería la primera versión del método: CRISP-DM Version 1.0.

El ciclo de vida de un proyecto de minería de datos, según el método CRISP-DM, pasa por seis fases. Dichas fases pueden verse en el esquema que presentamos, adaptado a nuestro propio proyecto, en la Figura 11. Las flechas muestran, resumidas, las secuencias y relaciones más destacables entre las fases (uni- o bi-direccionales).

Figura 11. Fases del ciclo de vida de un proyecto de minería de datos según el método CRISP-DM.



La secuencia de las fases no es rígida, al contrario, se requiere, con frecuencia, moverse en uno y otro sentido entre diferentes fases. El círculo exterior (de color azul)

simboliza la naturaleza cíclica de la minería de datos. El proceso no termina una vez que se ha llegado a la aplicación práctica del modelo. Al contrario, las experiencias acumuladas durante el proceso, así como las que aportan los resultados obtenidos con su aplicación, son fuentes de información que pueden permitir optimizar el modelo. Además el entorno en que se aplica es, normalmente, cambiante. Cambios en regulaciones y leyes, cambios socio-económicos, culturales, etc. pueden afectar al modelo y éste deberá ser revisado para adaptarlo a dichos cambios.

Para analizar los datos de nuestra muestra y llegar al momento de iniciar el desarrollo del modelo, hemos seguido la secuencia de tareas que muestra, esquematizada, la Figura 12.

Figura 12. Secuencia de tareas para el análisis de los datos.



Veamos, a continuación, de forma más detallada, los pasos que hemos seguido para la realización de esta investigación:

1. **Conocimiento de la IT.** A través de: la revisión de la bibliografía publicada, consulta con expertos, la puesta en común de experiencias propias, etc.
2. **Preparación de los datos.** Mediante: la recogida, extracción e integración de datos diversas fuentes para crear el almacén de datos; la validación y limpieza de los datos; el análisis y tratamiento de los datos perdidos; la realización de transformaciones de los datos y creación de nuevas variables; la reducción del número de categorías y de variables; etc.
3. **Exploración y conocimiento de los datos.** Realizamos: el Análisis Exploratorio de los Datos (AED); verificamos su calidad; y procedemos a describir los datos.

4. **Análisis de las Dependencias y asociaciones entre variables.** Se procedió a explorar las posibles asociaciones y dependencias entre los diferentes pares de variables.
5. **Exploración y selección de variables.** Analizamos qué variables aportan más información relevante para los objetivos del estudio y se seleccionan las más importantes para este fin. Realizamos también análisis de componentes principales. De este modo hemos podido reducir el número de las variables analizadas y utilizadas para crear los modelos. También eliminado, de este modo, el “ruido” que aportan las variables de menor importancia.
6. **Búsqueda y construcción de un modelo predictivo.** Hemos estudiado los patrones ocultos en los datos relacionados con la ITCC para poder conocer mejor cómo se comportan los diferentes factores en relación con la duración de la misma y poder diseñar algún modelo predictivo que facilite la gestión y la toma de decisiones con relación a ésta.
7. **Validación del modelo.** Después de haber diseñado y evaluado algunos modelos procedimos a valorar la bondad del modelo desarrollado cuando se generaliza su aplicación a poblaciones mayores. Esto es, validamos el modelo. Para ello utilizaremos una muestra (o varias) de pacientes para construir el modelo y otra diferente (u otras) para medir su riesgo de error. Hemos utilizado el método de validación cruzada con 10 submuestras (*10-folds*, según el término inglés). Este método divide la muestra en 10 subconjuntos y construye el modelo utilizando todas las submuestras menos una y lo prueba en la restante. Calcula el riesgo de error para ese modelo. Se repiten estos pasos de forma iterativa (10 veces) con cada uno de los subconjuntos. Finalmente, se obtiene un modelo final y se obtienen su riesgo de error calculando la media aritmética de los riesgos de error estimados en los 10 pasos previos.
8. **Aplicación del modelo a la práctica clínica diaria.** Una vez que se ha validado el modelo, para que sea de utilidad práctica, es necesario crear herramientas que permitan su aplicación. Estas herramientas generalmente consistirán en una aplicación informática que facilite la labor predictiva a partir de unos pocos datos. También pueden ser de otra naturaleza como, por ejemplo, un cuadro con reglas predictivas. En nuestro caso hemos procedido a crear: una escala de puntuaciones del riesgo de superar los estándares de Ibermutuamur, que hemos denominado con las siglas ERIT (Escala de Riesgo de la IT), y un cuadro de reglas predictivas. Éste último podría ser

transformado en algoritmos que permitan construir una aplicación informática que podría integrarse con la Historia Clínica electrónica.

9. **Adaptación recurrente del modelo.** Con posterioridad al inicio de su utilización en el día a día del control de la ITCC, será preciso confirmar su validez y adecuación a los fines perseguidos y, si fuera preciso, realizar los ajustes y modificaciones que fueren necesarios. Las características propias de la ITCC cambian con el tiempo, con la promulgación de nuevas normas que la regulen, con los cambios socio-económicos y culturales, etc. Esto obliga a revisar continuamente el modelo para adecuarlo a dichos cambios. Es decir, se trata de un modelo dinámico que debe evolucionar en el tiempo a la par que lo haga la ITCC.

Dependencias y asociaciones entre variables

Hemos analizado las posibles asociaciones y dependencias entre los diferentes pares de variables que tratamos en este estudio para conocer cómo se relacionan entre sí. Para consultar las relaciones y dependencias entre las diferentes variables consultar las siguientes tablas en los correspondientes anexos:

1. Para las asociaciones entre variables categóricas ver Anexo 4 (Matriz de los valores p de las pruebas Chi cuadrado entre variables categóricas).
2. Para pruebas ANOVA entre variables continuas y categóricas ver Anexo 5 (Matriz de valores p de los análisis de la varianza entre las variables continuas y categóricas).
3. Para las correlaciones entre variables continuas ver Anexo 6 (Tabla de correlaciones entre variables continuas).
4. Para las covarianzas ver Anexo 7 (Matriz de covarianzas).

En todas ellas podemos encontrar los valores p para las correspondientes pruebas (Chi cuadrado y ANOVA) o los valores de las mismas (correlaciones y covarianzas) para cada par de variables que se analizan.

Técnicas estadísticas utilizadas

En todos los casos se utilizó diversidad de técnicas estadísticas y gráficas en función de su utilidad para los objetivos planteados. Éstas incluyeron numerosas técnicas exploratorias y descriptivas de los variables y de sus relaciones como son: el cálculo de estadísticos descriptivos de centralidad y dispersión, de frecuencias absolutas y relativas, tablas de contingencia y cálculo del valor de X^2 , correlaciones, covarianzas, pruebas Anova, Manova, Ancova y MLG, pruebas no paramétricas, histogramas y gráficos de dispersión de puntos, gráficos categorizados, gráficos ternarios, de cajas y bigotes, etc. Se han utilizado también métodos de clasificación, árboles de decisión, reglas y otras técnicas empleadas en minería de datos para la búsqueda de patrones explicativos y/o predictivos. Entre las numerosas pruebas disponibles para el diseño de los modelos, hemos preferido el uso de algoritmos del tipo árboles de clasificación CHAID. Utilizamos como variables dependientes la duración de la ITCC para los modelos explicativos y, para los modelos predictivos, una variable dicotómica, creada *ad hoc*, que marca los casos que superan o no superan el estándar Ibermutuamur para cada uno de los diagnósticos. Como factores hemos tomado el resto de variables que seleccionamos previamente como más importantes por la información que aportan en relación con la ITCC.

Escala de puntuación de riesgo IT (ERIT)

Para facilitar la aplicación de los modelos a la práctica clínica diaria hemos creado una escala de puntuaciones que calcula el riesgo que tiene un paciente de superar el estándar de duración de la IT de Ibermutuamur para su diagnóstico. A dicha escala la denominamos *Escala de Riesgo de la IT* (ERIT). También creamos una tabla que resume dicha escala y las reglas que se deben seguir para calcular el riesgo correspondiente a cada paciente. Esta tabla permite clasificar a cada paciente y predecir la probabilidad que tiene de que su IT se alargue más allá del valor de referencia que es la duración estándar de Ibermutuamur para su diagnóstico.

Datos perdidos, selección de variables más relevantes y dependencias y asociaciones entre las mismas

A continuación veremos los métodos y resultados alcanzados con las técnicas que hemos empleado para:

1. Análisis de los datos perdidos.
2. Análisis de las dependencias, correlaciones y asociaciones entre las variables.
3. Selección de variables y reducción del nº de variables, casos y categorías.

Análisis de los datos perdidos

Los casos incompletos pueden ocasionar diversos inconvenientes al analizar estadísticamente los datos. Así, podremos encontrar patrones que se corresponden con pérdidas sistemáticas de datos y diferencias marcadas para una determinada variable entre grupos con datos perdidos y datos válidos para otra u otras variables. Esto podría conducirnos a obtener resultados engañosos. También puede producirse una pérdida en la precisión de ciertos estadísticos debido a la pérdida de información. Además muchos procedimientos estadísticos asumen (en sus planteamientos teóricos) que para realizar los cálculos contamos con casos completos. Por lo tanto, podríamos no cumplir dichas asunciones y desviarnos del desarrollo teórico en el que se basan los mismos.

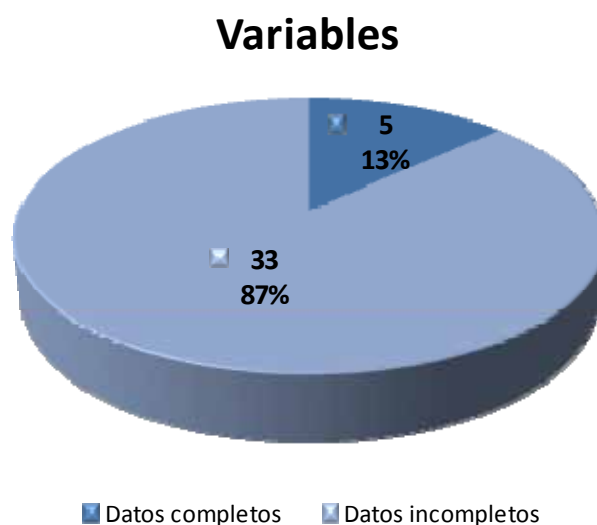
Es importante conocer la localización y extensión de los datos perdidos así como su relación con los valores de otras variables o de la propia. Y, en definitiva, si podemos afirmar que se han perdido al azar o, por el contrario, de forma sistemática y asociados a ciertos valores de alguna o algunas variables.

1. Localización y extensión de los datos perdidos

1.1. Variables con datos perdidos

En primer lugar observaremos el conjunto de variables que hemos escogido en la selección previa (es decir, las columnas del fichero de datos). Consideramos como variables completas aquellas que tienen datos para todos los casos (16.174 pacientes). Serán incompletas aquellas en las que falte un dato en un solo caso o varios datos en varios casos. Con este criterio, encontramos que tenemos un 13% de variables completas (ver Figura 13). A la hora de juzgar estos datos debemos de tener en cuenta que los criterios son muy exigentes y que para algunas variables determinadas puede no existir el dato para algunos pacientes de modo natural sin que ello signifique que el dato se “ha perdido”. Por ejemplo, un paciente que no ha sido intervenido no tendría fecha de intervención. Por lo tanto en algunas variables los “casos perdidos” no suponen un defecto sino que son consecuencia de la naturaleza propia de esa variable. Por ello consideramos que, en general, se ha producido una recogida de datos suficientemente adecuada y, como veremos en los siguientes análisis, mucho más completa de lo que pudiera parecerse a priori con este resumen por variables. En todo caso es incluso más importante analizar si dichas pérdidas se han producido al azar o, por el contrario, ha existido algún tipo de sesgo o de asociaciones entre diferentes variables que involucren a dichos valores perdidos.

Figura 13. Variables con datos perdidos.



En la siguiente tabla podemos ver el número de casos perdidos por variable (para las 30 variables más relevantes) ordenados por orden de frecuencia (ver Tabla 2).

Tabla 2: Tabla de datos perdidos

Variables	N Válidos	N Perdidos	%
Fecha de capacidad laboral	8728	7446	46,0%
Tipo de minusvalía	10329	5845	36,1%
Distancia hasta el trabajo	12345	3829	23,7%
Tipo de Contrato	12935	3239	20,0%
Antigüedad en años en la empresa	13131	3043	18,8%
Estudios	14421	1753	10,8%
Nacionalidad	14620	1554	9,6%
Médico titular o sustituto	14665	1509	9,3%
Estado civil	14793	1381	8,5%
¿Está pluriempleado?	14798	1376	8,5%
Edad en el momento de la baja (fechabaja-nacimiento)	14858	1316	8,1%
Sexo	14867	1307	8,1%
Fecha de nacimiento	14868	1306	8,1%
Actividad de la empresa	14872	1302	8,0%
Nº de hijos	14879	1295	8,0%
Ocupación	14890	1284	7,9%
¿Ha tenido IT en los 5 años previos?	14961	1213	7,5%
Fecha de la intervención	14972	1202	7,4%
Modo de pago	15253	921	5,7%
Provincia donde recibe la baja	15460	714	4,4%
Tipo de transporte	15536	638	3,9%
Base reguladora	15735	439	2,7%
Fecha del alta	16146	28	0,2%
Fecha de la baja	16164	10	0,1%
Diagnóstico principal	16167	7	0,0%
Rama de actividad	16174	0	0,0%
Nº de bajas	16174	0	0,0%
Causa del alta	16174	0	0,0%
¿Algún co-diagnóstico?	16174	0	0,0%
¿Le ha tratado el SPS, IB o ambas?	16174	0	0,0%

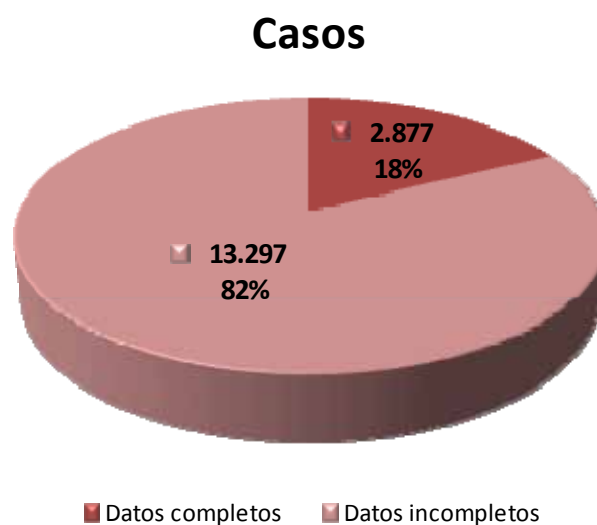
La variable objetivo de este estudio, la **duración de la IT**, es una variable calculada a partir de las fechas de alta y de baja. Para esta variable el número de casos perdidos es de 41, es decir el 0,25% de los casos. Por su escasa incidencia y

porque son datos perdidos completamente al azar, nos limitaremos a dejar fuera del estudio los mismos cuando corresponda.

1.2. Casos con datos perdidos

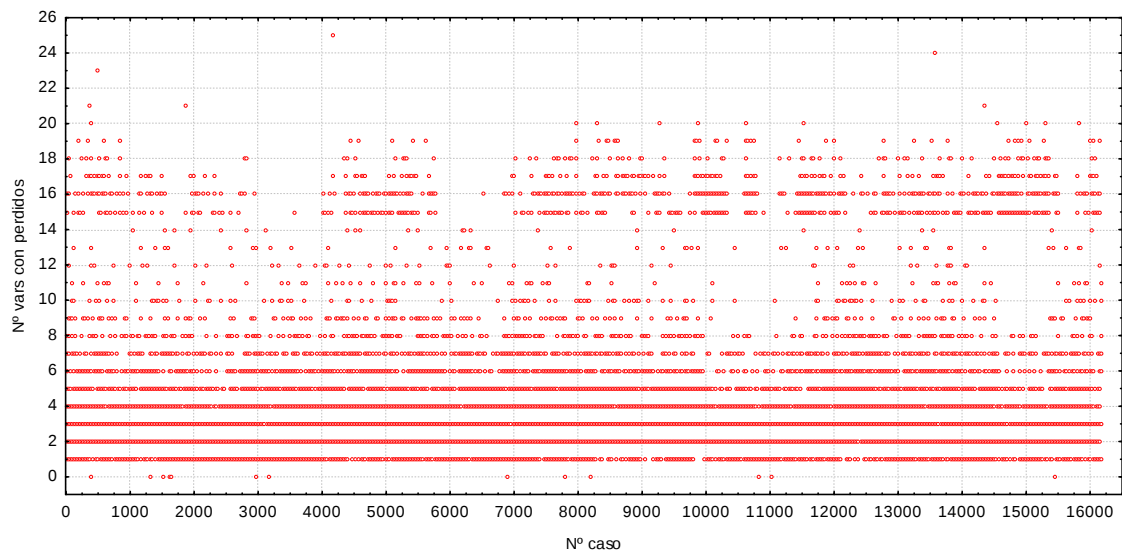
Al observar la base de datos por casos, es decir por filas de la tabla, encontramos un 18% de casos completos (ver Figura 14). Entendemos por casos incompletos aquellos en los que falta algún dato (o algunos datos) en alguna variable (o en varias).

Figura 14. Casos con datos perdidos



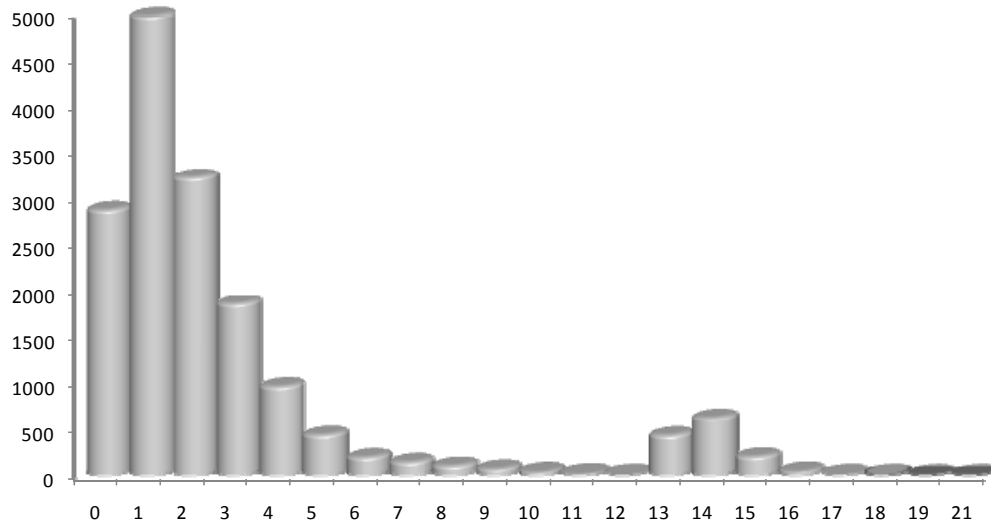
En la siguiente figura podemos contemplar para cada uno de los casos el nº de variables con algún dato perdido (Figura 15). Se han tenido en cuenta todas las variables y todos los casos contenidos en el fichero de datos.

Figura 15. Número de variables con algún caso perdido para cada paciente.



Visto de otro modo, encontramos que tenemos muchos casos que solamente presentan valores perdidos para 1, 2 o 3 variables (ver Figura 16).

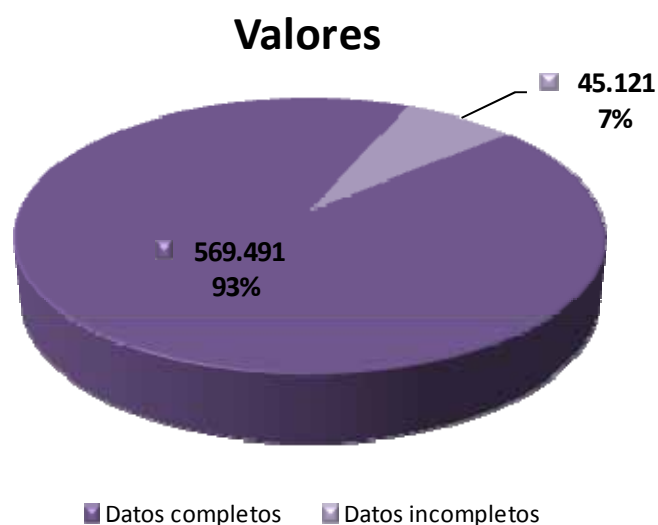
Figura 16. Histograma del número de variables con datos perdidos.



1.3. Datos perdidos por celda

Es por ello que al considerar los valores de forma individual, es decir cada celda de la tabla de datos, **tan solo se pierden un 7 % del conjunto todos los datos** (ver Figura 17). Con esta nueva perspectiva la base de datos aparece mucho más completa de lo que podía parecer al considerarlo por variables o por casos. Y mucho más si tenemos en cuenta que para algunas variables en natural que existan casos perdidos, como ya vimos al comienzo de este apartado.

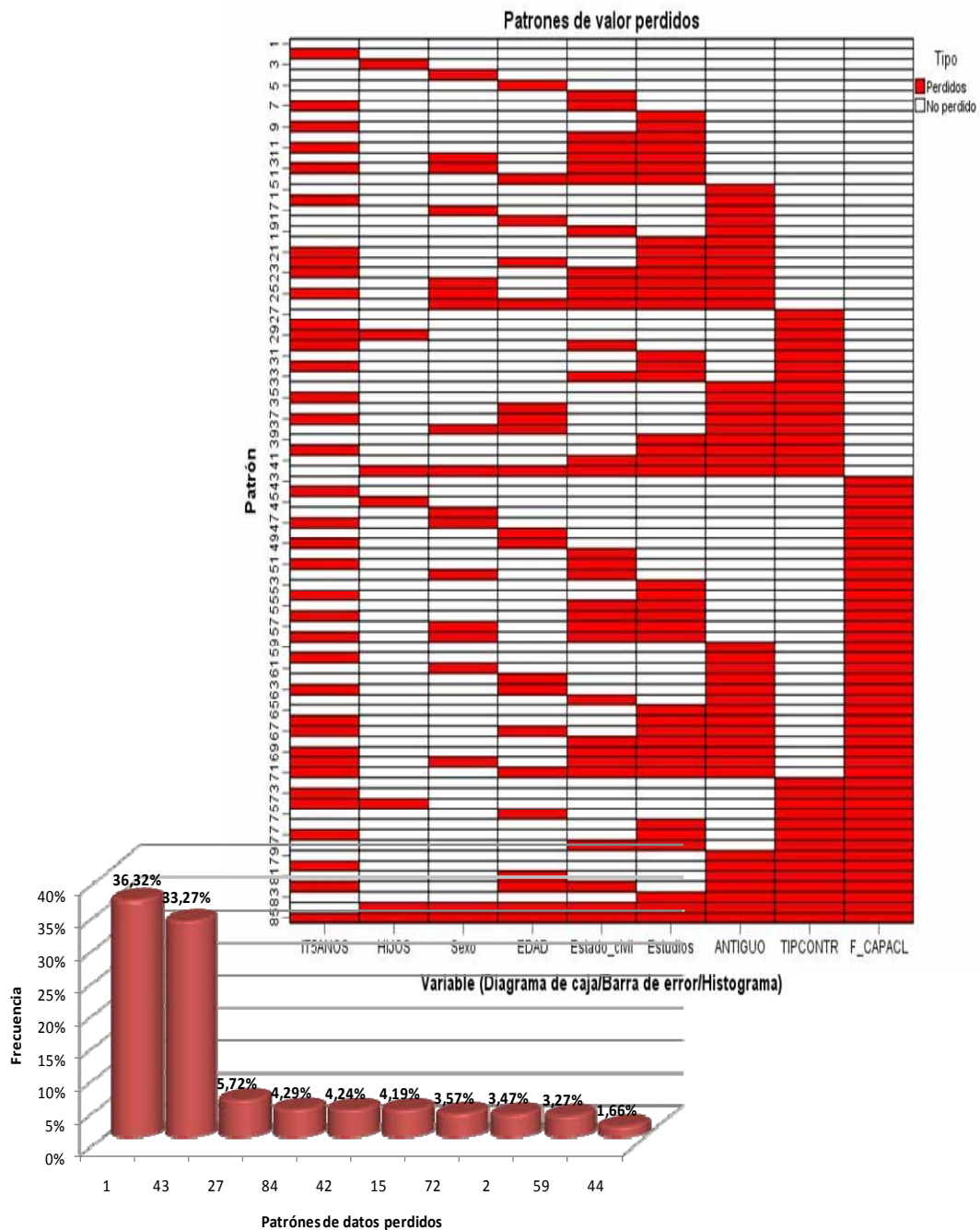
Figura 17. Celdas con datos perdidos.



2. Pérdidas de datos en las variables relevantes para este estudio

Vamos a seguir profundizando en el análisis comprobando los diferentes patrones de pérdidas de datos que se producen. Entendemos por patrones las diferentes combinaciones de variables con datos perdidos para cada caso concreto. Veremos un resumen de dichos patrones y la frecuencia con que ocurre cada uno de ellos en la Figura 18.

Figura 18. Patrones de valores perdidos y su frecuencia para las variables más relevantes.



Presentamos en la figura anterior (Figura 18) los patrones de datos perdidos para las variables más destacadas por este concepto. El patrón más frecuente es el 1 con más del 36% de casos. Dicho patrón corresponde a la situación en que no hay casos perdidos para ninguna variable. Como puede apreciarse dicho valor es muy diferente al 13% de la Figura 13. Las diferencias son debidas a que en la Figura 18 no estamos considerando todas las variables (como si hacemos en la Figura 13) ya que

hemos dejado fuera las variables que no van a ser relevantes para nuestro estudio. También hemos dejado fuera los patrones de datos perdidos que son muy poco frecuentes para facilitar la lectura del gráfico. De modo que las Figura 13 a Figura 17 nos proporcionan una visión global de cómo ha sido la recogida de los datos mientras que la Figura 18 nos da una idea de los patrones de datos perdidos para las variables consideradas como más relevantes para este estudio. El siguiente patrón más frecuente (nº 43) es el formado por los casos que tienen datos en todas las variables excepto en la correspondiente a la fecha de capacidad laboral, que es la variable con más datos perdidos (46%) como puede comprobarse en la Figura 18. El resto de patrones tienen frecuencias muy bajas (inferiores al 5%) o, incluso, suceden raramente. En la Figura 18 puede verse un gráfico con las frecuencias de los 10 patrones más relevantes. Hay que tener en cuenta, como ya hemos indicado, que para algunas variables es natural que existan casos perdidos ya que, por sus características, para algunos casos no puede existir un valor. Así, por ejemplo, puede no existir una fecha de intervención para los pacientes con diagnóstico no quirúrgico. Resumiendo, vemos que salvo el citado patrón 43, **apenas existen patrones que se repitan con una frecuencia llamativa**. En el siguiente apartado examinaremos más a fondo dicho patrón.

3. Relaciones entre los valores perdidos de diferentes variables.

Un aspecto importante con relación a los datos perdidos es conocer si existe algún tipo de asociación entre los valores perdidos de diferentes variables que pueda relacionarse con sesgos en la recogida de los datos. De hecho, cuando los datos se han perdido “completamente al azar”, es decir no existe relación entre los datos perdidos y los valores de alguna variable, el problema es fácil de resolver. Podemos dejar fuera del análisis los casos perdidos o sustituirlos por valores calculados a partir de los observados. Cuando los casos se han perdido “al azar”, es decir los casos perdidos en una variable se relacionan con determinados valores de otra u otras variables pero no con los valores propios, podemos utilizar la información que tenemos para tratar dichos casos perdidos y para conocer las causas probables que originaron dichas pérdidas. Si los casos perdidos se relacionan con los valores de la propia variable, como sucede cuando no tenemos la edad para los sujetos de un rango concreto de edades (si, por ejemplo, a los individuos de esa edad no les gustara desvelarla), no podremos solucionar los problemas e imprecisiones que se deriven de dicha pérdida.

Por ejemplo, si suelen perderse datos en una variable cuando se da cierto valor en otra variable o si la pérdida de datos en una variable independiente afecta de forma importante a los valores de la variable dependiente estudiada, en este caso, la duración de la IT. Sería el caso, por ejemplo, de que falte un dato como el sexo para los casos de una determinada provincia o de una ocupación concreta. Analizaremos éstas posibles asociaciones para conocer si los datos perdidos se distribuyen al azar y de forma independiente con relación a otras variables.

¿Qué relación hay entre los datos perdidos en “fecha de capacidad laboral” y los valores propios o de otras variables?

La situación más llamativa e importante se produce por la pérdida de datos en la variable que recoge la **fecha de capacidad laboral**. Ésta es la fecha que estima el médico de control de Ibermutuamur como adecuada para la reincorporación del paciente a su puesto de trabajo.

Al analizar los datos perdidos de la fecha de capacidad laboral no encontramos indicios de que pueda existir relación entre éstos y los valores de la propia variable, es decir con fechas o rangos de fechas determinados. En este sentido, podemos considerar que dentro de la propia variable se distribuyen al azar.

Con relación a otras variables encontramos:

1. Los datos perdidos no están relacionados con los valores de las siguientes variables: *edad, número de hijos, ocupación, nivel de estudios, nacionalidad, actividad económica, sector de actividad económica, trabajo de tipo manual o no manual, tipo de contrato, pluriempleo, antigüedad en el puesto de trabajo, base reguladora o número de bajas previas.*
2. Encontramos relaciones entre los datos perdidos de la fecha de capacidad laboral y algunos valores de las siguientes variables: *sexo, diagnóstico, co-diagnóstico, estado civil, provincias (y CC.AA.), modo de pago, causa de alta, derivación a especialista (por médico de AP), tipo de médico, número de días de bajas previas.*

En la Tabla 3 podemos ver resumidos las variables y sus correspondientes valores que están relacionados con una mayor o menor frecuencia de datos perdidos en la variable que estamos estudiando.

Tabla 3. Valores que modifican la frecuencia de datos perdidos en fecha de capacidad laboral

Variable		Valores
Mayor frecuencia	Sexo	mujer
	Tipo de médico	especialista
	Le atendió	médico de AP
	Co-diagnóstico	uno o más
	Causa de alta	Informe propuesta agotamiento de plazo otras causas
	Estado civil	viudo desconocido NS/NC
	Días en bajas previas	Mayor número de días
Menor frec.	Sexo	hombre
	Le atendió	especialista de SPS, IB o ambos
	Causa de alta	curación
	Modo de pago	Pago directo por cuenta propia

Algunas de estas variables y valores pueden actuar, en el asunto que nos ocupa, como meros marcadores o como variables de confusión. Por ejemplo, es posible que no sea el sexo del sujeto propiamente dicho sino otras variables relacionadas con el mismo, como por ejemplo el diagnóstico, las que realmente tengan una relación más estrecha con el origen de dichas pérdidas en los datos.

Otras variables cuyos valores se relacionan con los datos perdidos de *fecha de capacidad laboral* son: *diagnóstico* y *provincia*. A continuación veremos con más profundidad las mismas. En los siguientes gráficos podemos observar la proporción de datos perdidos con relación al total de casos para cada diagnóstico (Figura 19) y para cada provincia (Figura 20). La barra gris representa la frecuencia absoluta de casos recogidos con *fecha de capacidad laboral* para la correspondiente categoría. La barra de color salmón representa los casos sin dicha fecha para la misma. La suma de ambas será el total de los casos recogidos para la correspondiente categoría.

Figura 19. Valores perdidos en *fecha de capacidad laboral* y su relación con el diagnóstico.

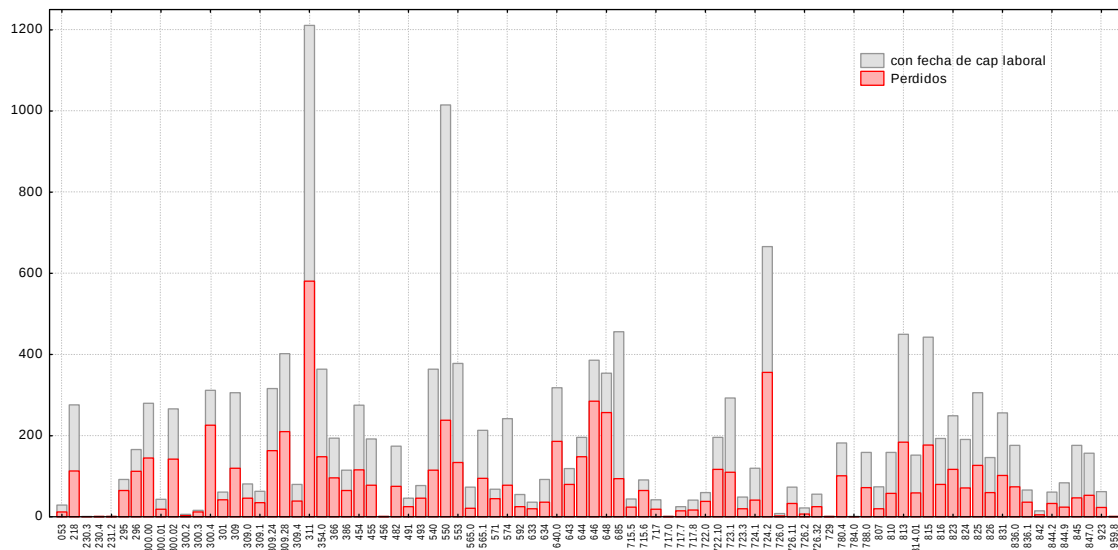
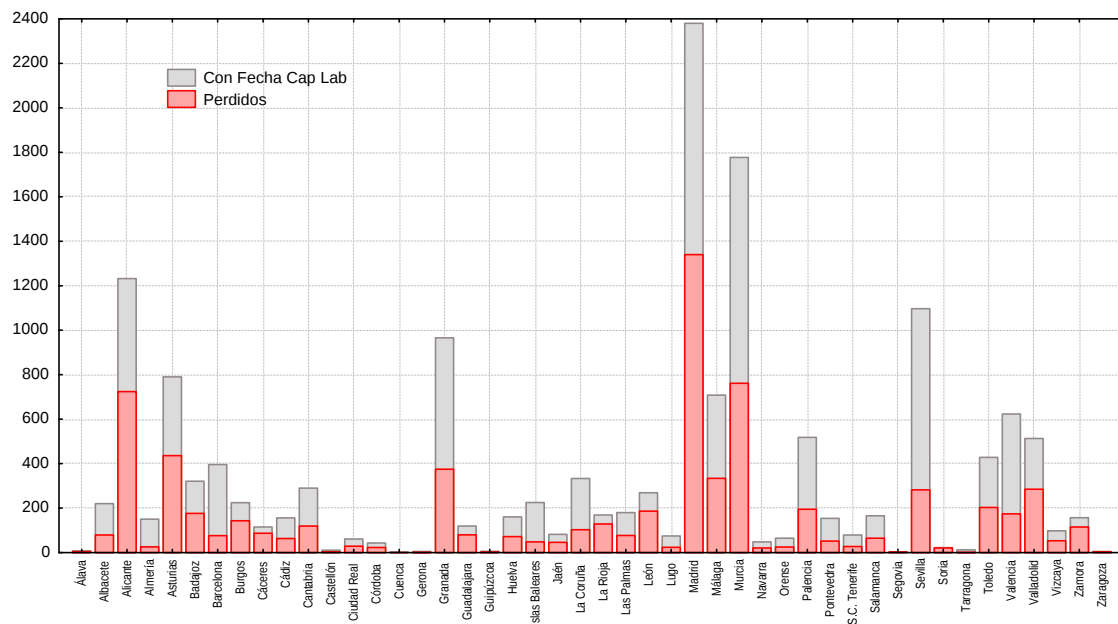


Figura 20. Valores perdidos en *fecha de capacidad laboral* y su relación con la provincia.



Podemos observar en ambos casos la disparidad de frecuencias relativas en lo que se refiere a los datos perdidos para la variable *fecha de capacidad laboral* que existe entre las diferentes categorías de las variables *diagnóstico* y *provincia*. Dicha disparidad puede apreciarse mejor en los siguientes gráficos. Éstos muestran la frecuencia relativa de datos perdidos en la citada variable para cada categoría de las variables *diagnóstico* (Figura 21) y *provincia* (Figura 22). La línea horizontal sirve como referencia y marca el nivel del 50%.

Figura 21. Proporción de valores perdidos en *fecha de capacidad laboral* por diagnóstico.

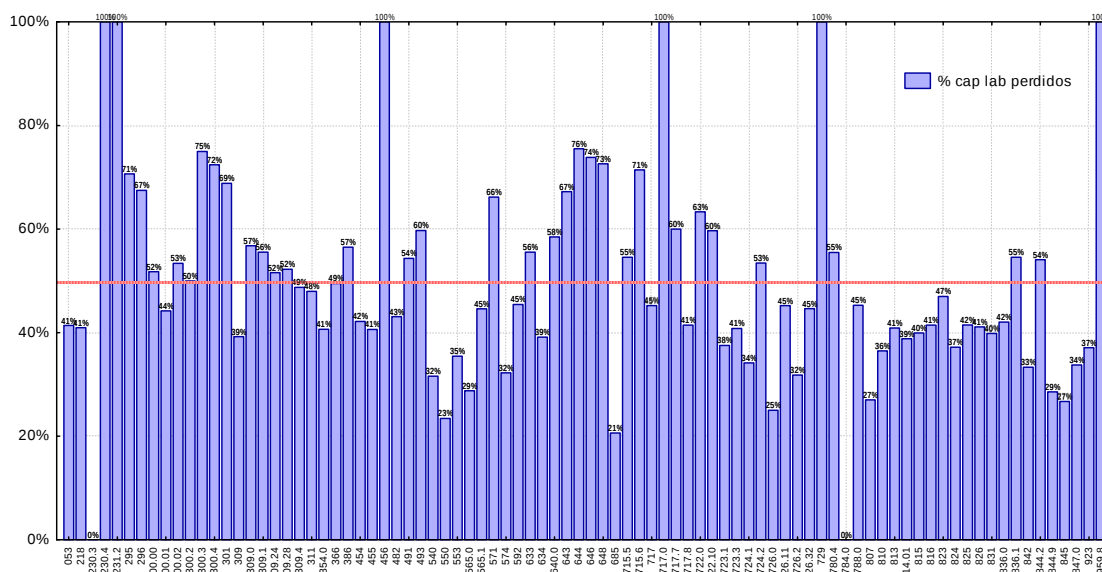
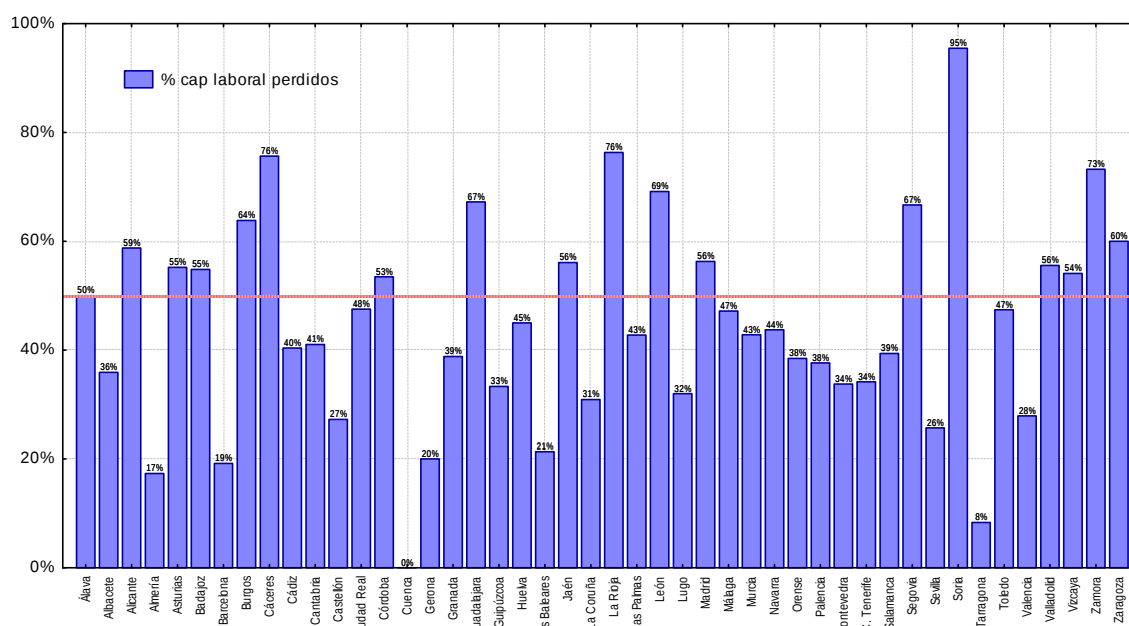


Figura 22. Proporción de valores perdidos en *fecha de capacidad laboral* por provincia.



Es evidente que hay una marcada relación entre los casos perdidos de la variable *fecha de capacidad laboral* y las diferentes categorías de estas variables. Es muy probable, incluso, que el hecho de presentar ciertos diagnósticos o de haber recibido la baja en determinadas provincias pueda estar relacionado directamente con el hecho de que el médico de control haya registrado la fecha en que estima que el paciente podría estar capacitado para su retorno al puesto de trabajo. Otras variables influirán también, como hemos visto en que se registre el citado dato. Así, por ejemplo, es más probable

que tengamos esa fecha si al paciente le ha visitado el especialista de Ibermutuamur, si no ha sufrido complicaciones en su proceso (presencia de codiagnósticos, agotamiento de plazo, informe propuesta, etc.), si se ha curado, etc.

Conclusiones en relación a los datos perdidos de la variable “fecha de capacidad laboral”

Según acabamos de ver, los datos perdidos en la variable “fecha de capacidad laboral” no se distribuyen “completamente al azar” sino que están relacionados con algunos valores de ciertas variables, anteriormente nombradas. Podemos considerar que no hay relación entre los mismos y los valores que pudiera tomar la propia variable, es decir, que dentro de dicha variable si se distribuyen al azar. En consecuencia, esta variable requerirá un tratamiento apropiado en aquellos análisis en que sea utilizada junto con otras variables. Dicho tratamiento es posible y recomendable en este caso. Cuando se utilice se explicará en el apartado en que se presenten los resultados del análisis correspondiente.

¿Cómo afectan los datos perdidos en la fecha de capacidad laboral a la variable duración de la IT?

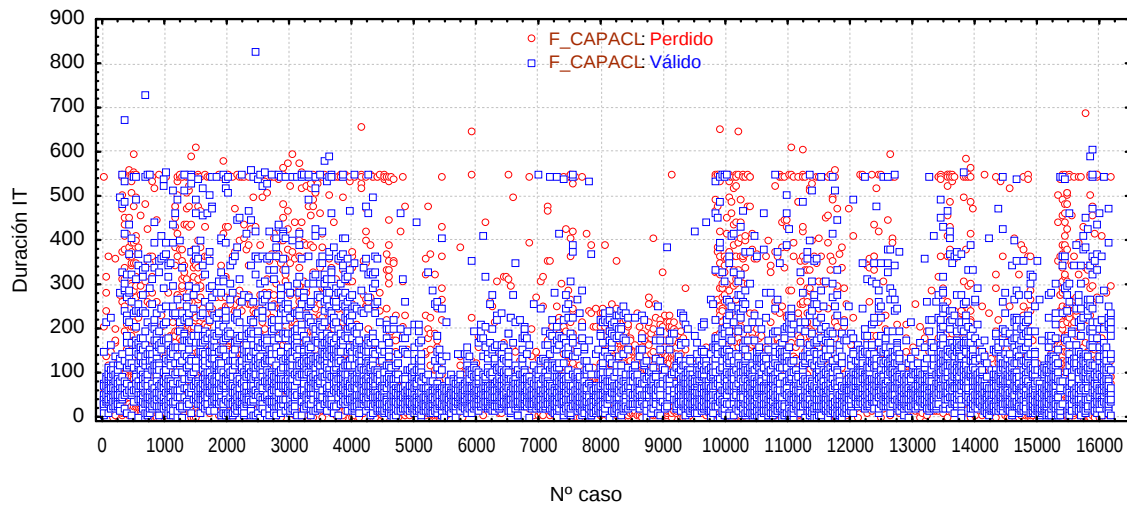
Realizamos pruebas T para grupos independientes con estimaciones separadas de las varianzas (da una mejor estimación con tamaños muestrales y varianzas dispares) utilizando la duración de la IT como variable dependiente y como variable de agrupación una variable indicador cuyos valores marcan qué casos tienen fecha de capacidad laboral y cuáles no la tienen. Con dicha prueba encontramos que la diferencia entre las medias de la duración de la IT correspondientes a los grupos formados por los casos válidos y los perdidos de dicha variable es estadísticamente significativa e igual a 32 días ($p < 0,01$), más elevada para los casos perdidos. Ver la siguiente tabla (Tabla 4).

Tabla 4: prueba T con estimaciones separadas de la varianza.

Duración IT	
Media perdidos	128,95
Media válidos	97,04
t-valor	17,02
gl	15247,00
p	0,00
t separ.var.est.	16,39
gl	11654,84
p 2-sided	0,00
Validos N perdido	6832
Validos N válido	8417
DE perdido	136,84
DE válido	93,75
F - Varianzas	2,13
p Varianzas	0,00
Diferencia entre medias	31,90
IC -95%	28,22
IC +95%	35,57

Como podemos observar en la tabla, la duración media de la IT es mayor para los datos perdidos de la variable *fecha de capacidad laboral*. También lo es la dispersión de sus valores (DE). Este resultado es congruente con otros resultados que ya hemos observado anteriormente como es el hecho de que aumente la frecuencia relativa de casos perdidos para dicha variable cuando el proceso se complica y alarga; como, por ejemplo, si existe comorbilidad, informe propuesta, agotamiento de plazo, ciertos diagnósticos (como el 311), etc.

Cuando observamos el diagrama de dispersión de puntos de la IT (para todos los casos del estudio) observamos que, aparentemente, los datos del grupo sin fecha de capacidad laboral se distribuyen al azar y no destaca un patrón muy diferente del que presentan los del grupo que si tiene dicha fecha (Ver Figura 23).

Figura 23. Valores perdidos y válidos para *fecha de capacidad laboral*.

Al comparar las distribuciones de frecuencias de la duración de la IT para ambos grupos (ver Figura 24 y Figura 25) podemos observar lo siguiente: para el grupo de casos con fecha de capacidad laboral hay una frecuencia ligeramente mayor en los rangos de duración de la IT cortos (por debajo de los 6 meses); las frecuencias tienden a ser similares para los dos grupos en los rangos medios y largos; y es mayor para el grupo de los que no tienen dicha fecha en el caso de aquellos que están en agotamiento de plazo. Esta observación puede explicar las diferencias encontradas en las medias. Sin embargo, parece que dichas diferencias afectan a un número de casos que sería, relativamente, no muy grande mientras que la gran mayoría de los mismos presentan una distribución similar para ambos grupos con relación a la duración de la IT.

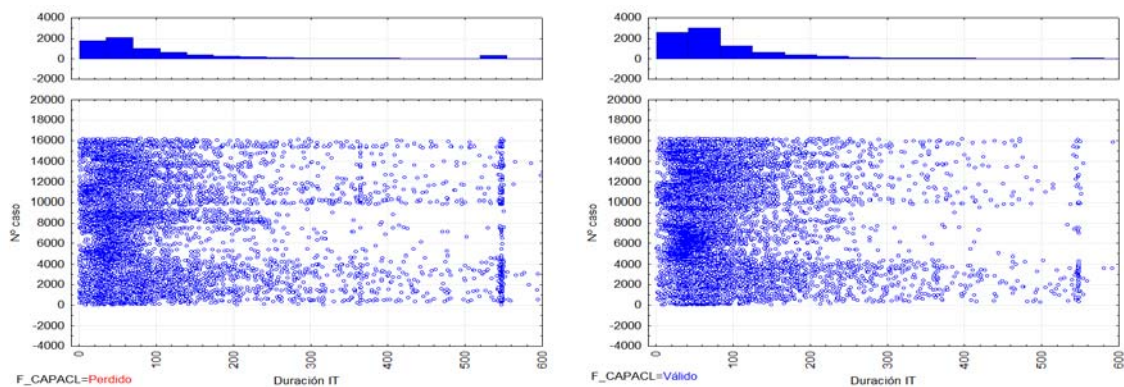
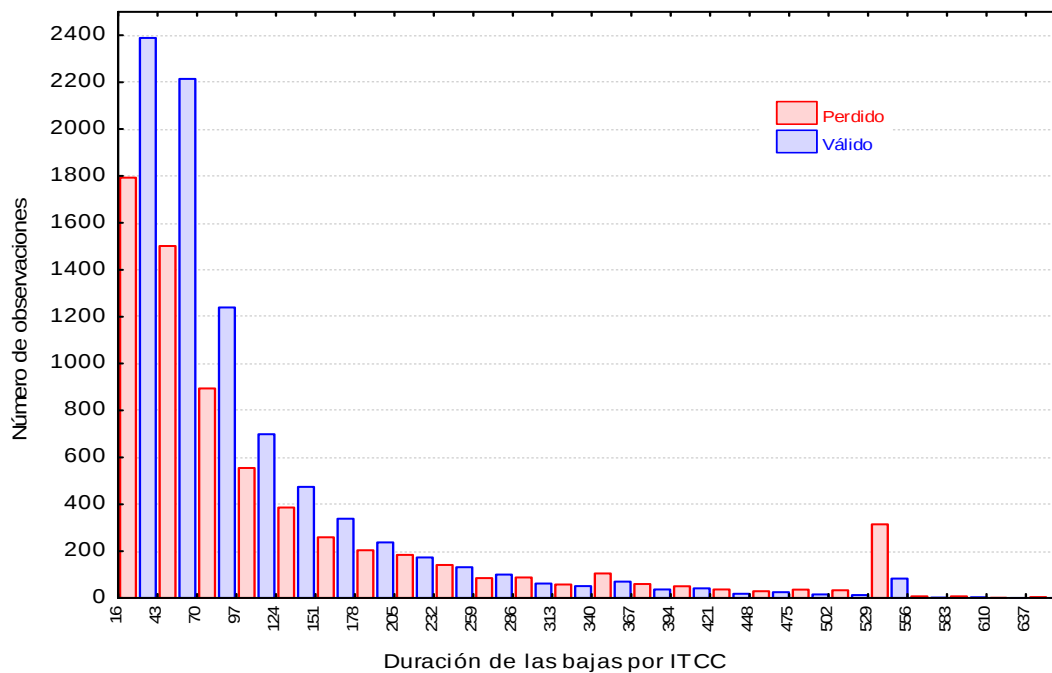
Figura 24. Comparación de los patrones de valores perdidos y válidos para *fecha de capacidad laboral*.

Figura 25. Relación de la duración de las bajas y los valores perdidos en *fecha de capacidad laboral*.



¿Qué relaciones encontramos entre los valores perdidos de las variables independientes y los valores de la duración de la IT?

Realizamos pruebas T para grupos independientes con estimaciones separadas de las varianzas porque ofrecen una mejor estimación cuando tenemos tamaños muestrales y varianzas muy dispares. Utilizaremos la duración de la IT como variable dependiente y como variables de agrupación indicadores cuyos valores marquen qué casos tienen valores perdidos o válidos para cada una de las variables independientes estudiadas (sexo, edad, provincia, etc.). Con ellas pretendemos conocer si existen diferencias significativas en la duración de la IT entre los grupos de casos con valores válidos y aquellos con valores perdidos para la variable independiente estudiada. Los resultados nos orientarán sobre la relación entre los valores perdidos de dichas variables y la duración de la IT.

Sexo

Hay diferencias estadísticamente significativas ($p < 0,01$) en las medias de la duración de la IT entre ambos grupos (con datos perdidos y con datos válidos); la

media de la duración es más alta para el grupo con los casos válidos (24 días de diferencia). Los casos perdidos muestran una distribución de frecuencias similar a la distribución propia de las duraciones de la IT; lo mismo se observa para el grupo con valores válidos. Dichas frecuencias son relativamente algo más elevadas en los valores de la duración más bajos para el grupo de los perdidos, lo que causaría las diferencias en las medias. Los valores aparentemente se han perdido al azar, con relación a la duración de la IT (ver Figura 26). Las diferencias en este caso parecen debidas, fundamentalmente, a una mayor frecuencia de casos marginales y extremos (y con valores más elevados) entre los valores válidos que entre los perdidos (ver Figura 27). Al analizar la duración de la IT por sexo podrían obtenerse duraciones medias algo más largas con relación a los valores esperados, pero no tenemos indicios de que esta desviación afecte de forma desigual a ambas categorías de *sexo*. Además solamente hay un 8% de valores perdidos para la variable *sexo*. Concluimos que las diferencias observadas, aunque estadísticamente significativas, se han producido al azar y debido, probablemente, a una mayor presencia de valores marginales en el grupo de datos no perdidos. No esperamos sesgos por la variable *sexo*.

Figura 26. Comparación de los patrones de valores perdidos y válidos para *sexo*.

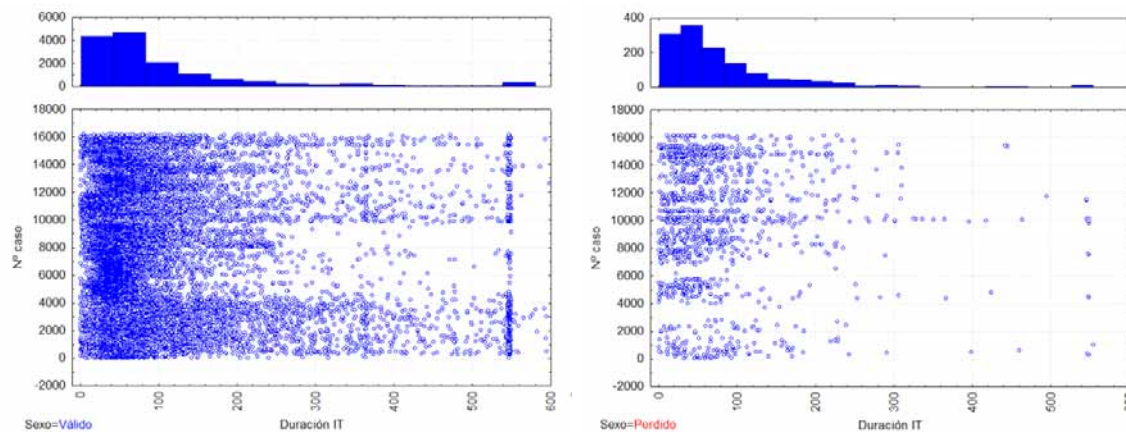
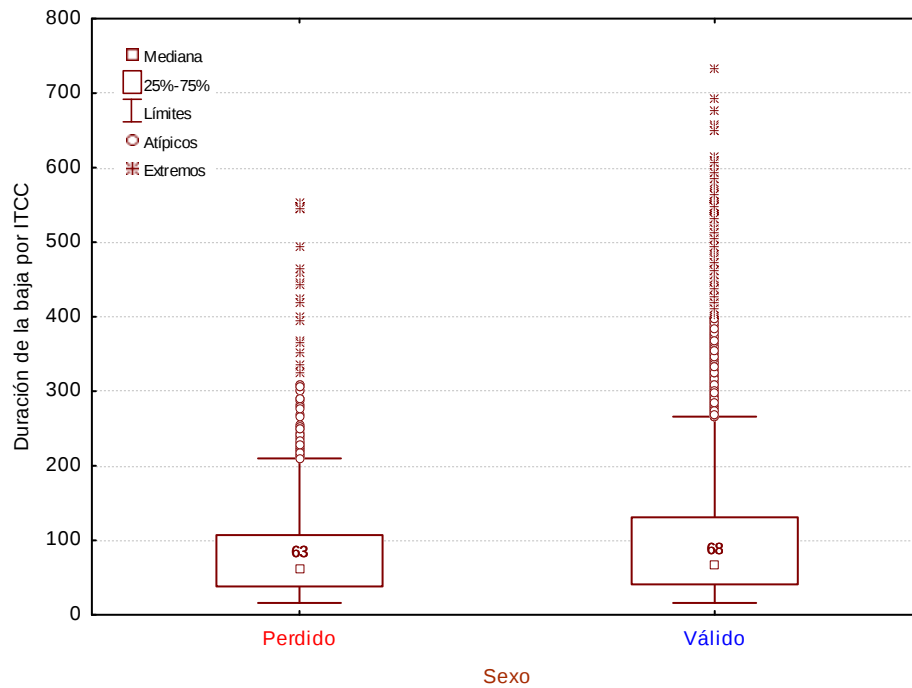


Figura 27. Distribuciones de la duración de la ITCC y su relación con los valores perdidos para la variable *sexo*.



Edad

Ocurre exactamente lo mismo que hemos descrito para la variable *sexo* (ver Figura 28 y Figura 29).

Figura 28. Comparación de los patrones de valores perdidos y válidos para *edad*.

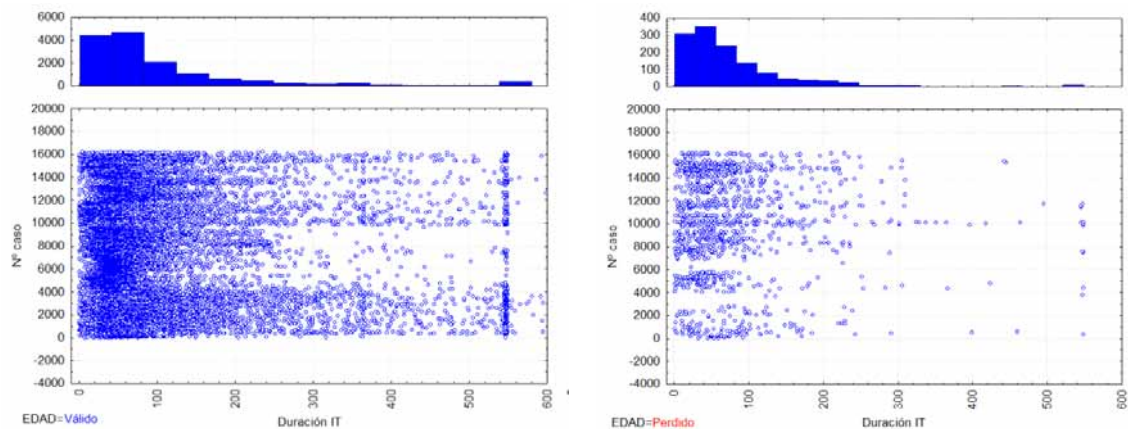
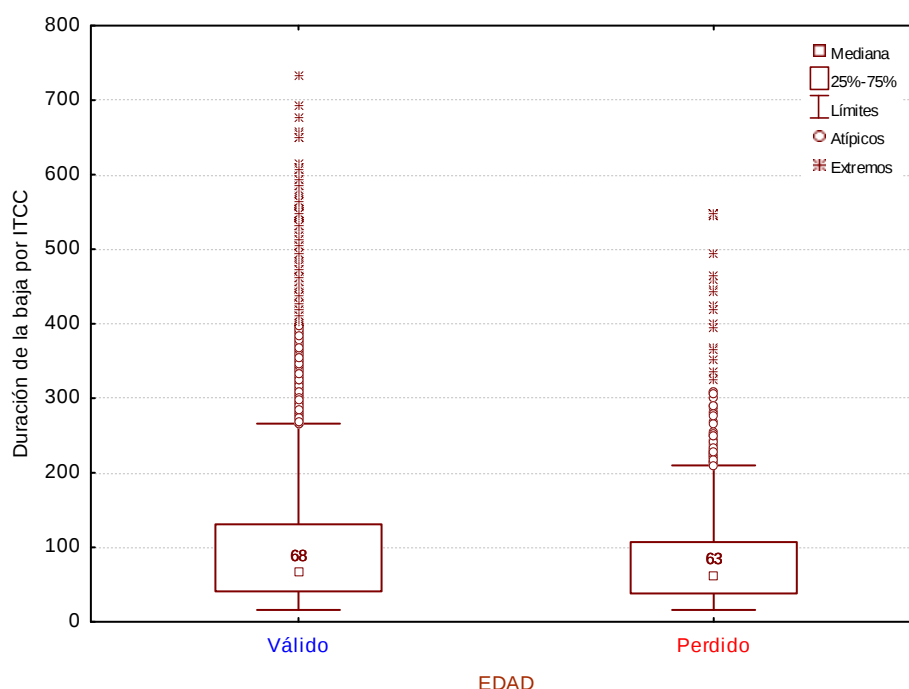


Figura 29. Distribuciones de la duración de la ITCC y su relación con los valores perdidos para la variable *edad*.



Esto es debido a que prácticamente se pierden datos de ambas variables en los mismos casos (con muy pocas excepciones, solo 51 casos; ver Tabla 5). La diferencia es la misma (24 días; $p < 0,01$) y también a favor de los casos válidos.

Tabla 5. Relación entre los valores perdidos de *sexo* y *edad*.

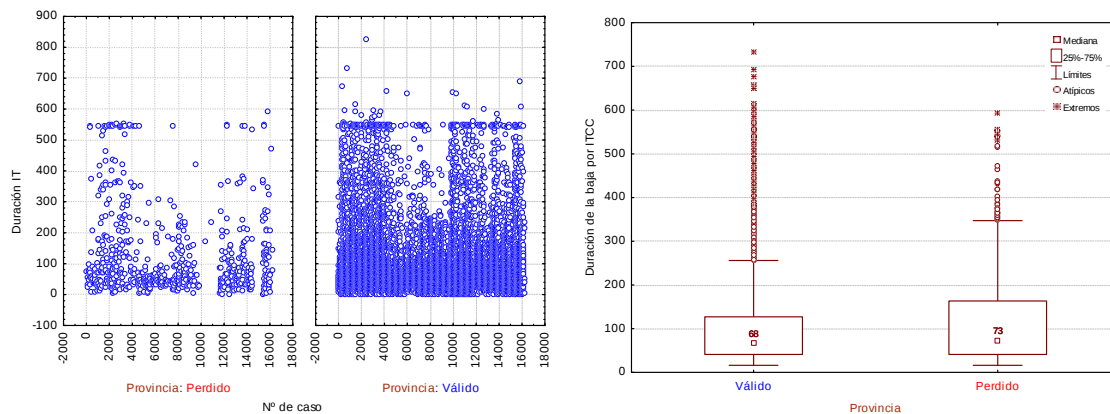
		Edad		Total
		Perdidos	Válidos	
Sexo	Perdido	1.286	21	1.307
	Válido	30	14.837	14.867
Total		1.316	14.858	16.174

Provincia.

La proporción de datos perdidos es del 4,41%. La distribución de la duración de la IT tiende a ser similar en ambos subgrupos (de datos válidos y de datos perdidos) y no parece seguir ningún patrón definido (ver Figura 30). En este caso la media de la duración de la IT es más elevada en el grupo con casos perdidos (23 días de diferencia

entre ambos). La diferencia entre ambos grupos es estadísticamente significativa ($p < 0,01$).

Figura 30. Comparación entre los patrones de valores perdidos y válidos para provincia y de las distribuciones para las duraciones de la IT para ambos subgrupos.



Actividad económica de la empresa y ocupación.

Encontramos diferencias estadísticamente significativas de 28 días ($p < 0,01$) en las duraciones medias de la IT para los grupos de casos válidos y perdidos para ambas variables (actividad y ocupación). Aunque la duración de la IT tiende a ser mayor en el grupo con valores válidos, los casos perdidos no parecen seguir ningún patrón concreto. Los valores perdidos se distribuyen aparentemente al azar con relación a la duración de la IT para ambas variables. Los valores perdidos de ambas variables coinciden en los mismos casos (salvo en 2 pacientes; que suponen el 0,001% del total de la muestra).

Otras variables relevantes.

Para otras variables, como son: *número de hijos, estado civil, nivel de estudios, nacionalidad, base reguladora, tipo de pago, tipo de contrato, tipo de transporte y pluriempleo*; también hay diferencias estadísticamente significativas entre los grupos con valores válidos y perdidos en las medias de la duración de la IT. Sin embargo, con frecuencia, se trata de diferencias bastante más pequeñas y, además, observamos que los valores perdidos se distribuyen al azar. Para la variable que recoge los años de antigüedad en el puesto de trabajo no se observaron diferencias significativas entre ambos grupos.

¿Qué relaciones encontramos entre los valores perdidos de las variables independientes?

Al analizar las asociaciones entre los valores perdidos y válidos de las variables sexo y edad con los correspondientes al resto de las variables (comparándolas dos a dos) encontramos solo pequeñas discrepancias. Dichas discrepancias son algo mayores para variables como *provincia* y aquellas que recogen datos de tipo administrativo como son: *antigüedad en el puesto de trabajo*, *base reguladora*, *tipo de contrato*, *tipo de pago* y *tipo de transporte*.

4. Conclusiones del análisis de datos perdidos

En el conjunto de la base de datos del estudio no encontramos una cantidad de datos perdidos excesiva (7%). Y además, como hemos visto, hay pocos patrones de datos perdidos que se repitan con gran frecuencia (ver Figura 18) para las variables más relevantes. Sin embargo si existe una pérdida importante de datos en el caso de unas pocas variables y especialmente para *fecha de capacidad laboral* (46% de casos perdidos). Los datos perdidos en la variable *fecha de capacidad laboral* están relacionados con algunos valores de ciertas variables. Por lo tanto, ésta variable requerirá un tratamiento apropiado en aquellos análisis en que sea utilizada junto con las mismas. Para el resto de variables, hemos observado que, en general, los datos perdidos se distribuyen al azar y no encontramos indicios objetivos para creer que existan sesgos importantes por esta causa, si bien conviene ser cautelosos y tenerlos presentes al realizar los correspondientes análisis estadísticos. Cuando desarrollemos los modelos predictivos emplearemos técnicas que clasificarán los datos perdidos como una categoría más y, por lo tanto, no deberán preocuparnos más que las consecuencias que se deriven de la interpretación de los correspondientes resultados.

Dependencias y asociaciones entre variables.

Hemos analizado las posibles asociaciones y dependencias entre los diferentes pares de variables que tratamos en este estudio para conocer cómo se relacionan entre sí. Para estudiar las relaciones y dependencias entre las diferentes variables se

pueden consultar las siguientes tablas en los correspondientes anexos de este documento:

1. Para las asociaciones entre variables categóricas ver Anexo 4. Matriz de los valores p de las pruebas Chi cuadrado entre variables categóricas.
2. Para pruebas ANOVA entre variables continuas y categóricas ver Anexo 5. Matriz de valores p de los análisis de la varianza entre las variables continuas y categóricas.
3. Para las correlaciones entre variables continuas ver Anexo 6. Tabla de correlaciones entre variables continuas.
4. Para las covarianzas ver Anexo 7. Matriz de covarianzas.

En todas ellas podremos encontrar los valores p para las correspondientes pruebas (Chi cuadrado y ANOVA) o los valores de las mismas (correlaciones y covarianzas) para cada par de variables que se analizan.

Selección de variables y reducción del nº de variables, casos y categorías

En este estudio se han recogido un gran número de variables y, algunas de ellas, contienen, además, gran cantidad de categorías diferentes. Si bien es conveniente contar con la mayor cantidad de información disponible y que la misma sea lo más detallada que sea posible, a la hora de realizar los análisis estadísticos de los datos tanta información es, a veces, difícil de manejar y, sobre todo, puede ser complicado entenderla e interpretarla. Por otro lado no toda la información disponible es igual de relevante para los propósitos del estudio. Una parte de dicha información, aunque fuera válida, correcta e interesante en sí misma, puede afectar tan solo a un grupo pequeño de los casos; o bien puede presentar efectos que sean débiles o poco relevantes para los objetivos del estudio; o, incluso, que dichos efectos solo sucedan en determinadas circunstancias y, por lo tanto, sean de menor interés al tratar de generalizar los resultados obtenidos. Por todas las razones que acabamos de exponer es preciso realizar una selección previa de las variables que vamos a tomar en consideración para analizar tanto la IT como los factores que influyen sobre la misma, así como para la construcción de un modelo predictivo que facilite su gestión. También

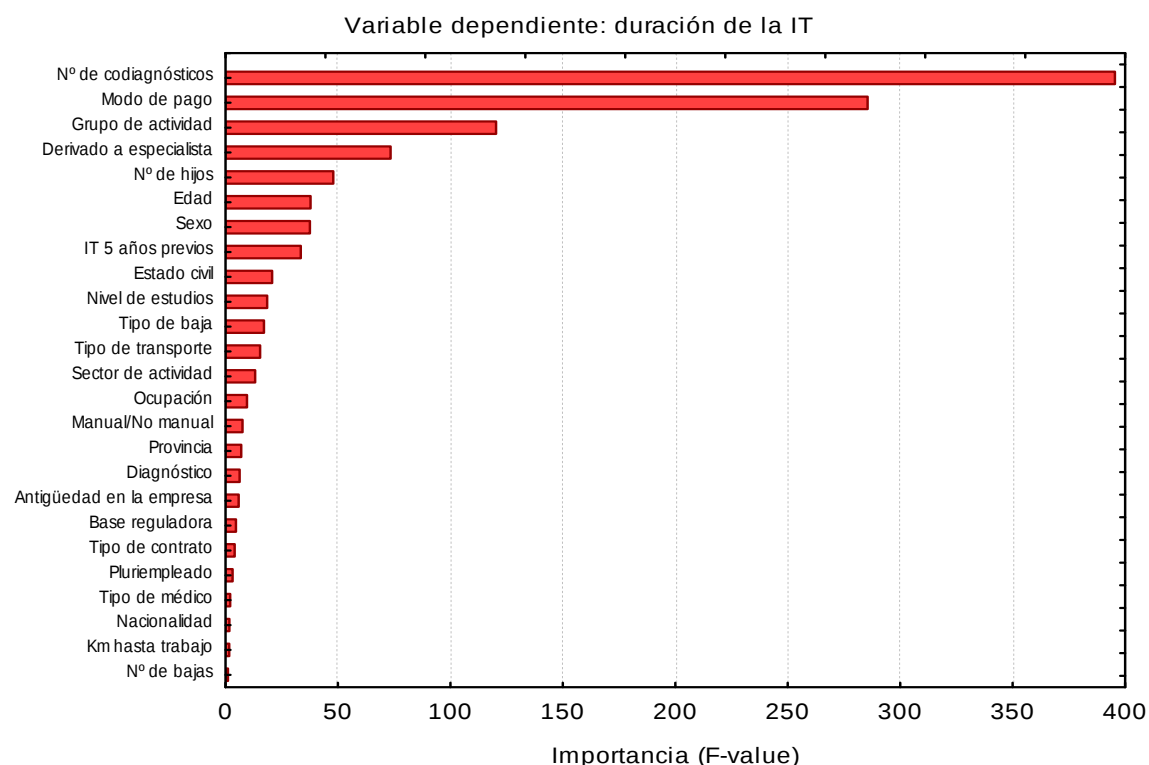
nos veremos obligados a reducir el número de categorías en aquellas variables seleccionadas que presentan demasiadas. El resto de la información y de los efectos propios de las variables no seleccionadas, aunque pudieran ser en sí mismos válidos, para los fines de este estudio los consideraremos como “ruido”; es decir, como información de escasa utilidad que, en cambio, pudiera desorientar o confundir al investigador.

Exploración y selección de las variables que aportan más información útil

Exploramos en primer lugar qué variables pueden presentar efectos más acusados sobre la IT. Tomamos como variable dependiente la variable *duración de la IT*. Dejaremos fuera de este análisis la variable *diagnóstico*, ya que sabemos de antemano que por su naturaleza influye directamente en la duración de la IT y puede hacer menos visible el efecto que tienen el resto de las variables. Además, la mayor parte de los análisis estadísticos los vamos a realizar posteriormente para un grupo de diagnósticos seleccionados previamente. También dejaremos fuera la variable *causa del alta* porque, si bien tiene una relación muy marcada con la duración de la IT, no se trata de un factor que actúe previamente influyendo en la variabilidad de la misma; ni tampoco es una variable con valor predictivo, ya que sus valores se establecen al finalizar la IT, cuando el médico extiende el alta; y, sin embargo, si que puede distorsionar los resultados que buscamos.

En la Figura 31 podemos ver una clasificación de las variables analizadas por orden de importancia de su efecto sobre la duración de la IT. En dicha figura, la variable denominada “Grupo de actividad” se corresponde con la reducción que hemos realizado de las numerosas categorías que contiene la variable *actividad económica* de la empresa agrupándolas en 7 nuevas categorías, como veremos un poco más adelante (ver página 66).

Figura 31. Importancia de los factores por su efecto sobre la duración de las bajas por ITCC.



Reducción del nº de categorías de las variables Diagnóstico y Actividad económica

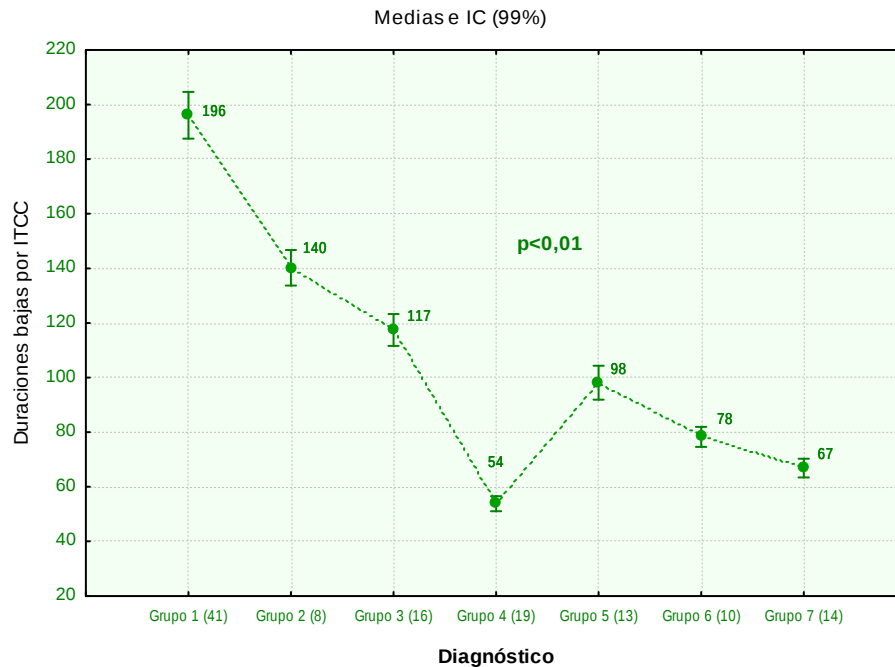
Algunas variables, como son *diagnóstico* y *actividad económica de la empresa*, contienen muchas categorías; más de 600 cada una de ellas. Un número tan grande de categorías no permite realizar correctamente los análisis y, además, sería muy difícil interpretar los resultados. Por ello procedimos a reducir su número mediante una técnica que consiste en agrupar aquellas categorías que son similares entre sí en lo que se refiere a su efecto sobre la duración de la IT; las categorías pertenecientes a un grupo son, todas ellas, suficientemente diferentes de las categorías clasificadas en los demás grupos en relación con dicho efecto.

Diagnósticos.

Podemos clasificar el conjunto de los diagnósticos en 7 grupos cuyos elementos presentan características próximas entre sí en lo que a la duración de la ITCC se refiere (ver Anexo 8). Entre las medias de la duración de la IT de cada uno de dichos grupos si hay diferencias estadísticamente significativas ($p < 0,01$ para todas las comparaciones posibles) como podemos apreciar en la Figura 32. Las cifras que aparecen entre

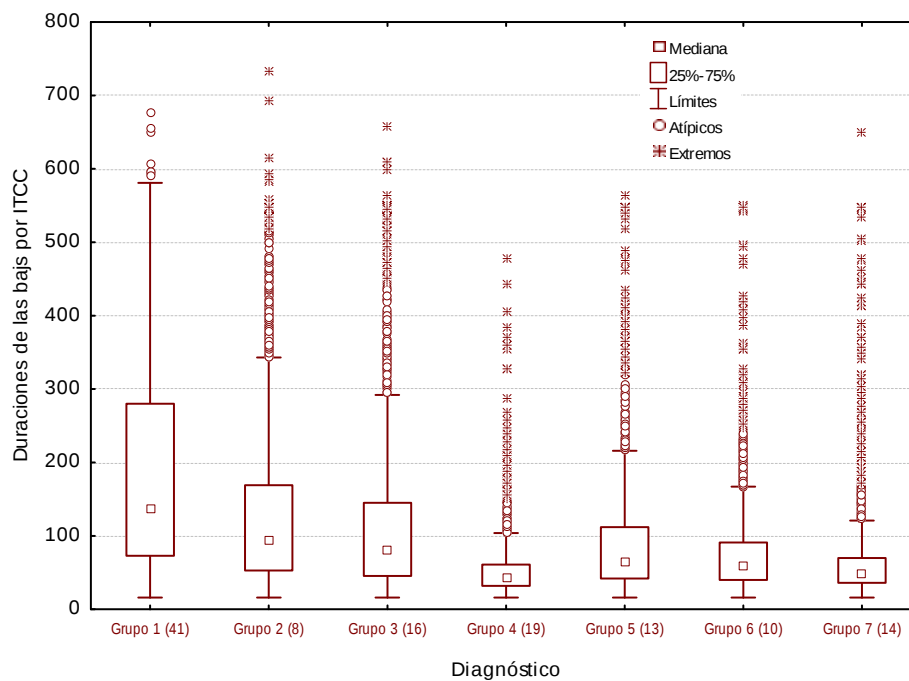
paréntesis después del nombre de cada grupo se corresponden con el número de categorías que contiene dicho grupo.

Figura 32. Duraciones medias de las bajas por grupo diagnóstico.



También son diferentes las distribuciones de la duración de la IT (Figura 33).

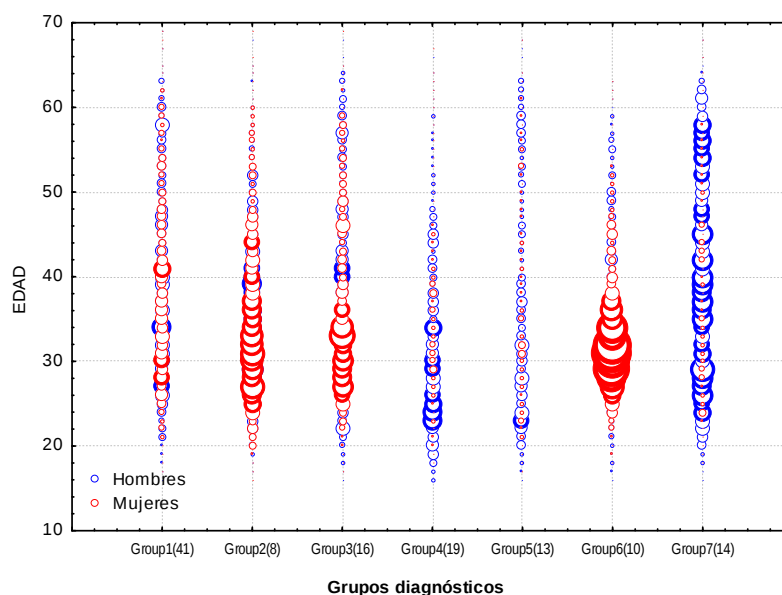
Figura 33. Distribuciones de las duraciones de las bajas por grupo diagnóstico.



En la tabla del Anexo 8 (Tabla con los diagnósticos, CIE-9, agrupados), pueden consultarse los diferentes diagnósticos recogidos en cada uno de los mencionados grupos.

Observamos, además, que hay marcadas diferencias en las distribuciones de frecuencias por sexo, edad y duración de la IT entre los diferentes grupos. Podemos apreciar las distribuciones para sexo y edad en la Figura 34. El diámetro de la burbuja está relacionado con la frecuencia de casos que se corresponde con cada punto que marca la intersección entre un grupo y una edad determinados. El color indica el sexo de los sujetos representados por la burbuja. Así, por ejemplo, vemos que hay una elevada frecuencia de mujeres en edades medias (25 a 40 años) en el grupo 6. O que en el grupo 7 predominan los hombres aunque con edades que se dispersan a lo largo de todo el rango de la variable.

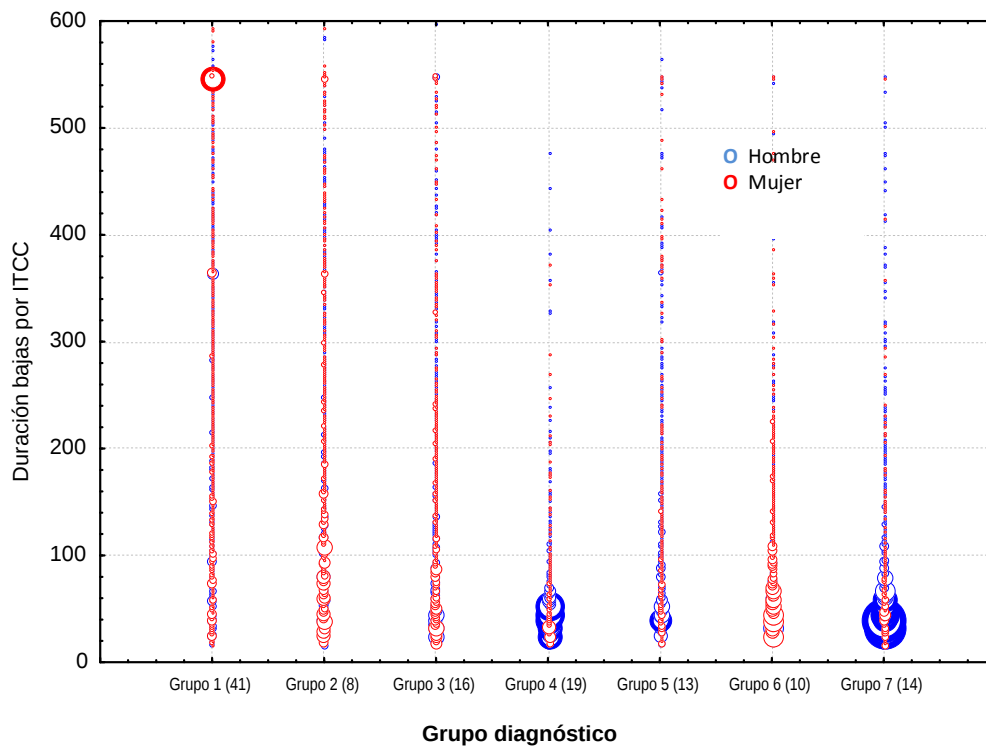
Figura 34. Frecuencias por edad y grupo diagnóstico.



En la Figura 35 podemos ver la distribución por grupo, sexo y duración de la IT. Así, por ejemplo, podemos observar cómo en los grupos 4 y 7 predominan los hombres con duraciones más bien cortas de la IT; mientras que en el grupo 6 predominan mujeres con duraciones cortas. Por el contrario, en el grupo 1 se encuentran, aparentemente con frecuencias similares, sujetos de ambos sexos y con duraciones de la IT que se distribuyen a lo largo de todo el rango posible de valores; aunque

podemos observar en este grupo un predominio de sujetos que se encuentran en agotamiento de plazo.

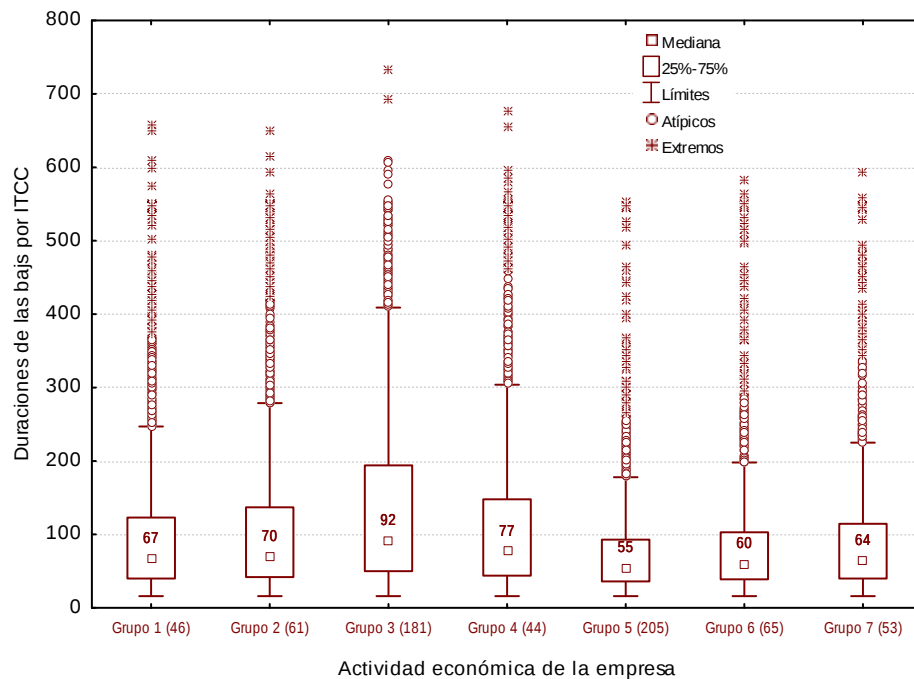
Figura 35. Frecuencias de las duraciones de las bajas por grupo diagnóstico y sexo.



Actividades económicas (según CNAE)

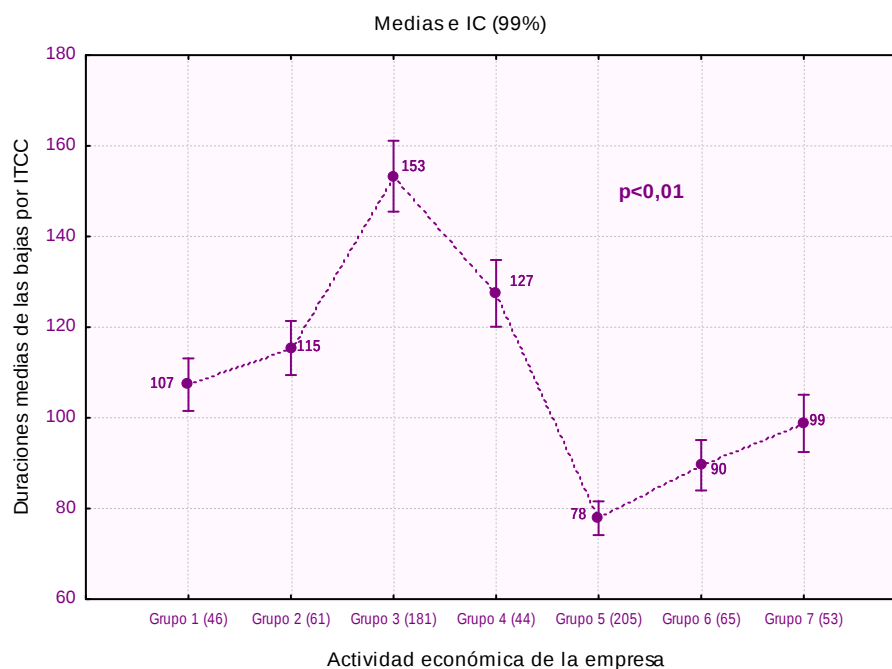
Las categorías correspondientes a la variable *actividad económica de la empresa* las hemos agrupado también en 7 grupos. Puede consultarse en la tabla del Anexo 9 (Actividades económicas de la empresa, CNAE, agrupadas). la relación de actividades económicas incluidas en cada grupo. En la Figura 36 podemos ver las distribuciones representadas por las medianas y cuartiles de cada uno de los 7 grupos.

Figura 36. Distribuciones de las duraciones de las bajas por grupo de actividad económica.



En la Figura 37 se representan las medias de la duración de la IT y sus intervalos de confianza al 95% para cada uno de los grupos. Hay diferencias estadísticamente significativas ($p < 0,01$) entre todas las medias de los diferentes grupos tanto cuando las comparamos entre si, como cuando consideramos el modelo en su conjunto. Siendo más exigentes no estarían tan claras las diferencias entre los grupos 1 y 2, el 6 y 7 y entre los grupos 1 y 7.

Figura 37. Duraciones medias de las bajas por grupo de actividad económica.



Limitaciones y posibles sesgos del estudio

Las conclusiones del estudio solo podrán aplicarse con garantías a sectores de la población general similares a la población laboral protegida por Ibermutuamur en relación con la contingencia común y a su implantación tanto territorial como por sectores y ocupaciones laborales. Extrapolar los resultados de este estudio a la población general o a otros grupos de la población puede llevar a conclusiones sesgadas o incorrectas. Antes de hacerlo sería preciso comparar dichos resultados con los de otros estudios realizados bien en la población general o bien en otros subgrupos de población diferentes del utilizado en este estudio.

La muestra se ha recogido de forma secuencial según acudían los pacientes a la consulta hasta completar un máximo de 20 pacientes por día y por médico. Por lo tanto, no se hizo al azar (siguiendo estrictamente la terminología de las técnicas de muestreo); aunque asumimos que no puede haber más sesgos que los propios que puedan presentarse si comparamos la población objetivo del estudio con la población general, como veremos a continuación; es decir, podemos considerar que se ha realizado el muestreo “al azar” siguiendo el sentido más popular del término. La

muestra fue recogida por todo el territorio español en función de la presencia de Ibermutuamur en cada zona del mismo. La representación de las diferentes zonas, por lo tanto, no tiene por qué ser homogénea ni en función de su población general. También la representatividad por sectores laborales y ocupaciones depende de las empresas que tienen estos servicios contratados con Ibermutuamur. Por lo tanto tampoco representará fielmente a todas las ocupaciones y sectores de la población general española. Por último, la representación por edades y sexos no se corresponde con las proporciones que podemos encontrar en la población general sino con las de la población objetivo de este estudio.

Por último, los modelos predictivos se han construido para los valores recogidos en un periodo temporal concreto. Diversas condiciones de todo tipo (laborales, normativas, socio-económicas, culturales, etc.) puede influir, y de hecho lo hacen, sobre el diseño y utilidad de los mismos. Dichas condiciones evolucionan y cambian con el paso del tiempo. Por ello, mismos precisarán de ajustes y modificaciones que los adecuen a la situación actual y a las venideras. Esto es algo natural y propio de estos sistemas y de las técnicas de data-mining. El punto fuerte de los modelos que se presentarán en el presente estudio no se encuentra tanto en los resultados propiamente dichos como en el método utilizado y las herramientas creadas con este fin. Éstos pueden ser reproducidos y actualizados continuamente para que las aplicaciones se adecuen a las situaciones cambiantes cada momento en el futuro.

Calendario previsto para el estudio

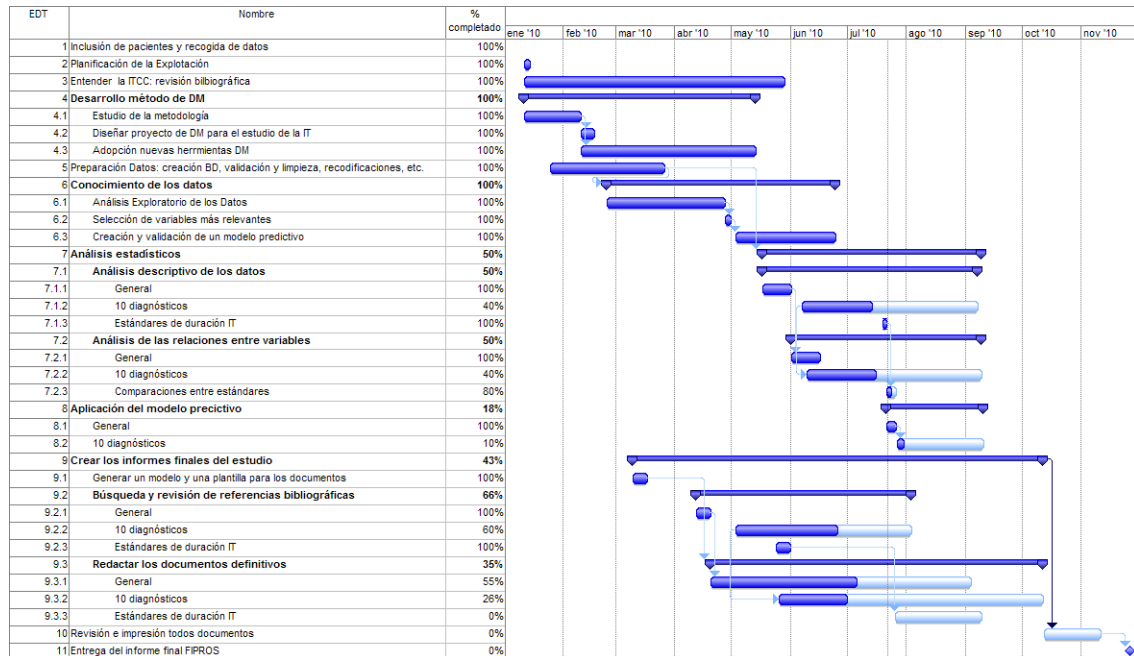
La recogida de los datos se realizó hasta completar la muestra siguiendo el calendario previsto entre enero del año 2003 y marzo del año 2007. El análisis de los datos y la búsqueda de modelos predictivos, objeto del presente estudio, se realizaron entre los meses de enero y noviembre de 2010.

Plan de trabajo

Se creó un plan de trabajo detallado con la ayuda de herramientas informáticas como Mindject MindManager 8 y Microsoft Office Project 2007. Dichas herramientas sirvieron posteriormente para realizar el seguimiento de la progresión del proyecto. En la Figura 38 podemos ver un resumen de las principales tareas planificadas junto a los

periodos de tiempo asignados a las mismas, así como el seguimiento de su estado de realización tal y como se encontraba a finales del mes de julio.

Figura 38. Plan de trabajo.



Cumplimiento de las exigencias éticas y legales

Se cumplió toda la normativa vigente en España con relación a las implicaciones éticas, legales y de protección de datos personales de los pacientes. El estudio carecía de riesgos para la salud de los pacientes por lo que no fue necesario tomar ninguna medida al respecto.

A cambio puede aportar importantes beneficios en lo que se refiere a mejorar nuestros conocimientos sobre la ITCC lo que, a su vez, podría aportar mejoras en el seguimiento y control de la misma y en la optimización de los recursos y del gasto relacionado con ella. También podría darnos pistas para ofrecer un mejor servicio a los pacientes que precisan una baja por contingencia común.

Resultados del estudio

En este capítulo presentaremos los resultados de este estudio ordenados de acuerdo con el siguiente esquema:

1. Análisis descriptivo de las variables más relevantes: demográficas, laborales, clínicas y administrativas.
2. Análisis de las relaciones entre dichas variables.
3. Análisis de la duración y del coste de la ITCC. Descripción de los componentes de la ITCC (tramos de duración de la IT).
4. Comparación entre dos diagnósticos relevantes: 311 y 550.
5. Búsqueda de patrones ocultos en los datos. Diseño y validación de un modelo predictivo.
6. Aplicación del modelo predictivo. Escala de Riesgo de la IT (ERIT).

Descripción de las variables y análisis de la relación entre las mismas

Variables Socio-demográficas

1. Sexo y edad

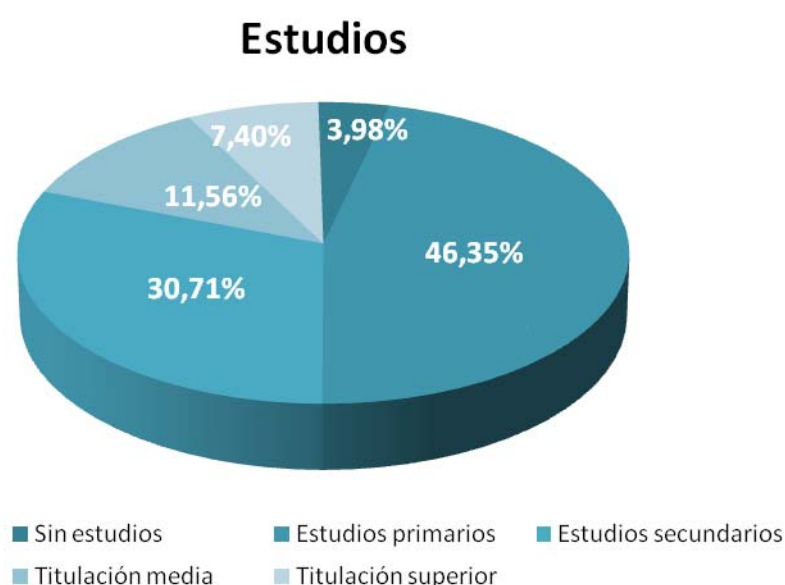
La descripción de las variables **sexo** y **edad** se ha visto anteriormente en el apartado “Edad y sexo de los pacientes incluidos en el estudio”, en la página 18, como descripción de la muestra de pacientes utilizada en el estudio. Como veíamos entonces, en nuestra muestra hay un 42% de mujeres y un 58% de hombres con edades medias de 37 años para las primeras y de 39 para los segundos.

A continuación presentamos la descripción de otras variables demográficas relevantes del estudio, a saber: *nivel de estudios, nacionalidad, estado civil, número de hijos y provincia donde recibió la baja*.

2. Nivel de estudios

En nuestra muestra casi el 4% de los trabajadores no tienen estudios. La gran mayoría de los trabajadores tienen estudios primarios (46,35%) o secundarios (30,71%). El resto o bien tienen titulaciones medias (11,56%) o bien superiores (7,40%). En general, son más frecuentes, en nuestra muestra, los niveles de formación más bajos (Figura 39).

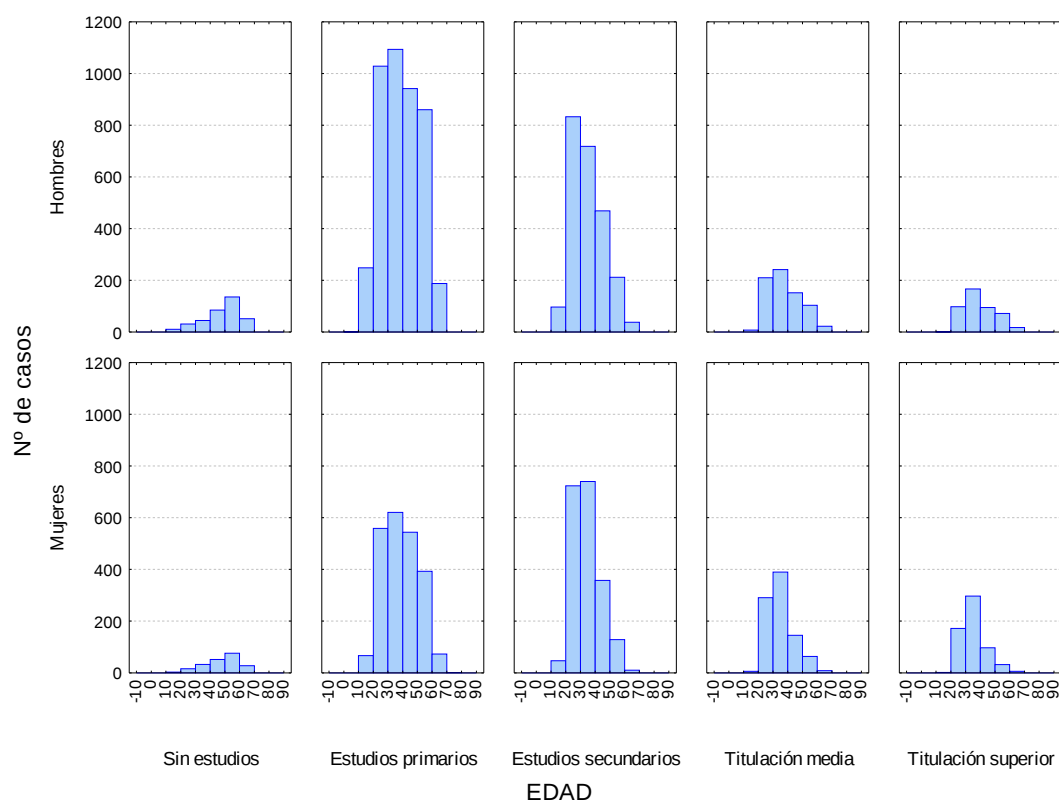
Figura 39. Frecuencias relativas de los niveles de estudio.



Estudios por edad y sexo

Las distribuciones por edad y nivel de estudios son bastante similares para ambos sexos; aunque podemos apreciar una tendencia hacia niveles de estudios más elevados entre las mujeres jóvenes y de edad media comparándolas con los hombres de edades similares. En la Figura 40 podemos observar cómo para las mujeres las frecuencias más elevadas se encuentran en el nivel de estudios secundarios y en edades comprendidas entre los 20 y los 40 años. Para hombres las frecuencias más elevadas se encuentran en el nivel de estudios primarios para edades entre los 20 y los 60 años. Además en los niveles correspondientes a las titulaciones medias y superiores las frecuencias absolutas son mayores para las mujeres entre los 20 y los 40 años que para los hombres en edades similares.

Figura 40. Histograma del nivel de estudios por edad categorizado por sexo.

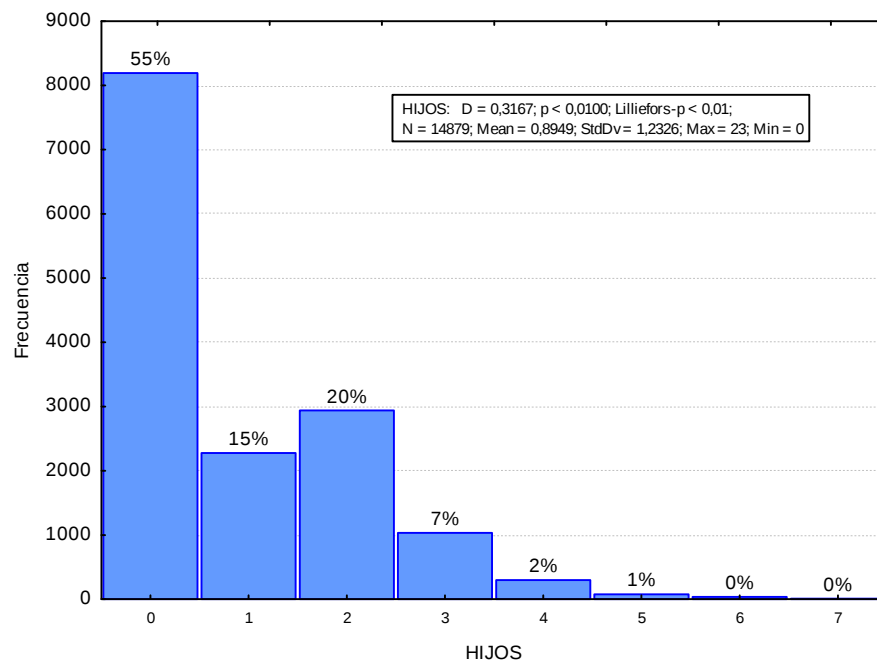


Parece evidente que en nuestra muestra **las mujeres jóvenes o de edad media tienen, en general, un mayor nivel educativo que los hombres de edades similares.**

3. Número de hijos

Observamos que más de la mitad (55%) de los trabajadores de la muestra no tienen hijos. Un 20% de los casos tienen 2 hijos y un 15% solo uno. A partir de tres la frecuencia cae vertiginosamente. Ver el histograma de la Figura 41.

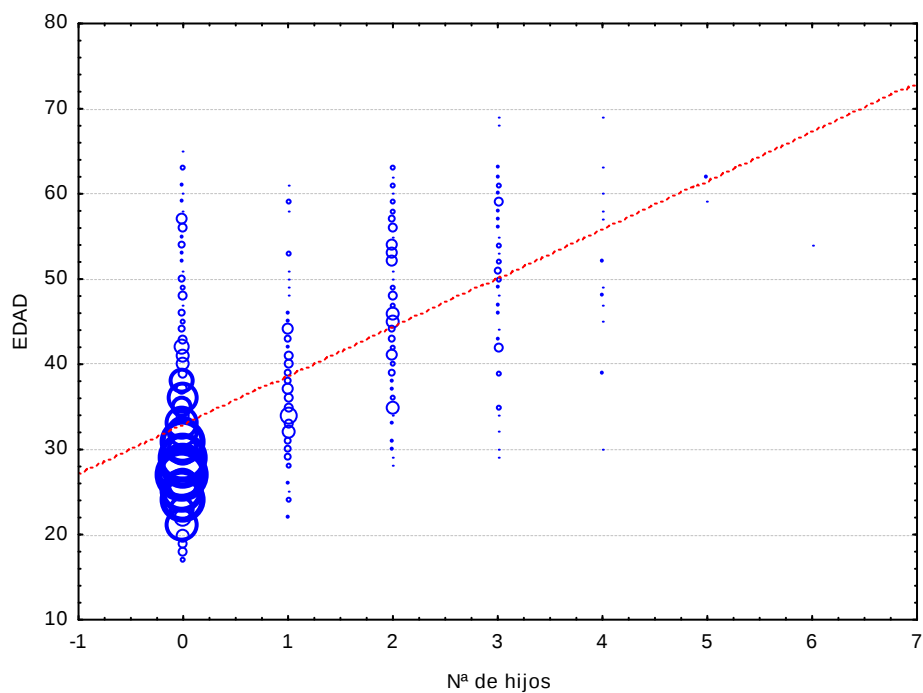
Figura 41. Histograma del número de hijos.



Nº de hijos por edad y sexo.

Como podemos apreciar en el diagrama de burbujas de la Figura 42, el número de hijos tiende a aumentar con la edad, aunque la gran mayoría (55%) de los trabajadores de nuestra muestra no tiene hijos; éstos tienen entre 20 y 40 años.

Figura 42. Diagrama del número de hijos por edad.

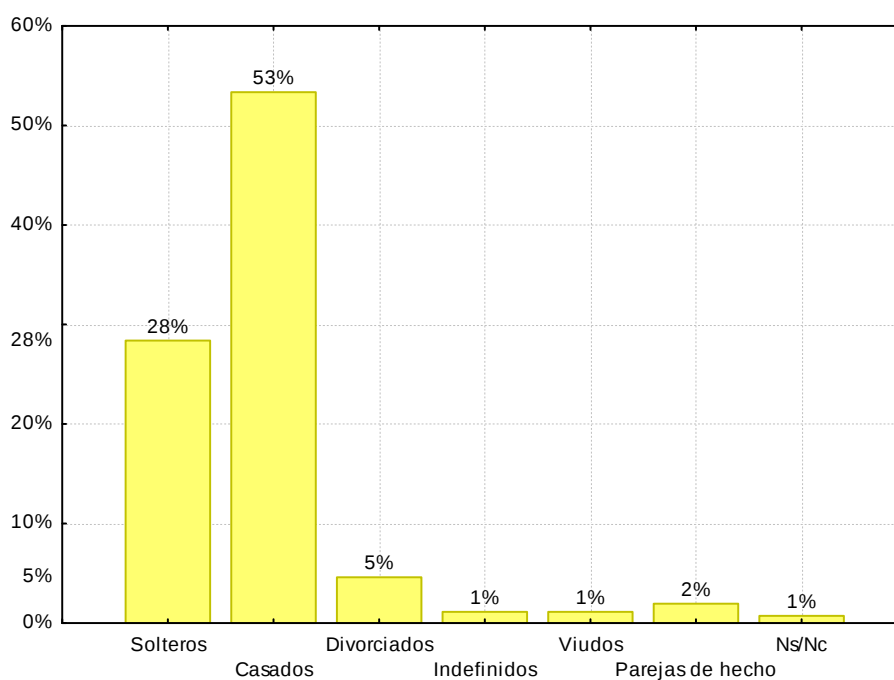


Encontramos, por lo tanto, una marcada tendencia, entre los trabajadores de nuestra muestra, a tener pocos hijos. Entre los trabajadores más jóvenes (20 a 30 años) es muy poco frecuente tener hijos (solo un 13% los tiene). Aumenta dicha frecuencia entre los 35 y los 60 años (casi el 67% tiene hijos) y, además, tiende a aumentar el número de hijos con la edad. De hecho existe una correlación positiva ($r=0,515$) y estadísticamente significativa ($p<0,01$) entre ambas variables.

4. Estado civil

Una mayoría de los sujetos de la muestra están casados (53%) o solteros (28%). El resto de opciones son muy poco frecuentes en nuestra muestra. Ver Figura 43.

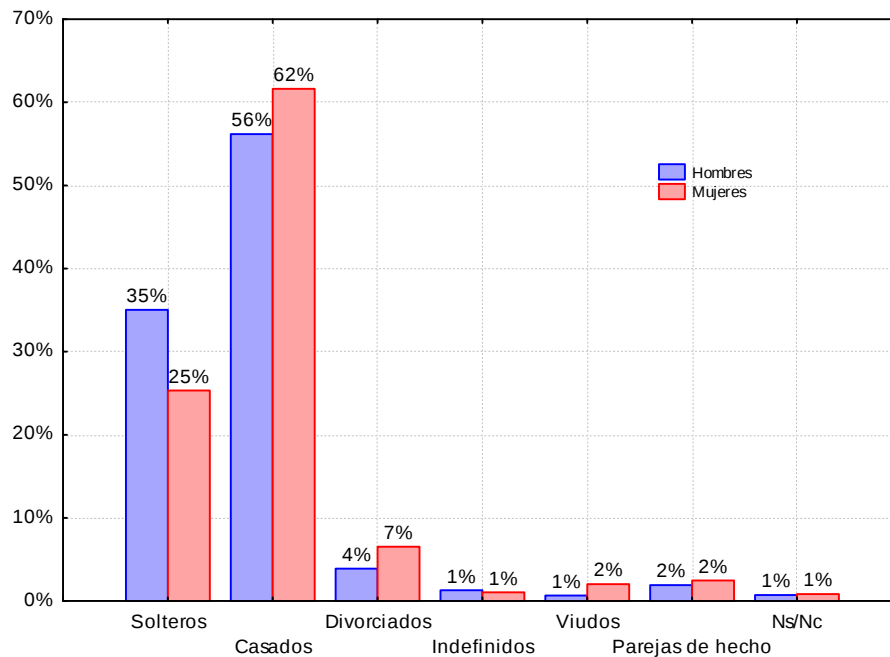
Figura 43. Histograma del estado civil.



Estado civil por sexo y edad

Entre las mujeres de nuestra muestra hay una mayor frecuencia relativa de casadas que entre los hombres (62% vs 56%) y de divorciadas (7% vs 4%). Entre éstos es más elevada la frecuencia de solteros (35% vs 25%). Consultar el diagrama de barras de la Figura 44.

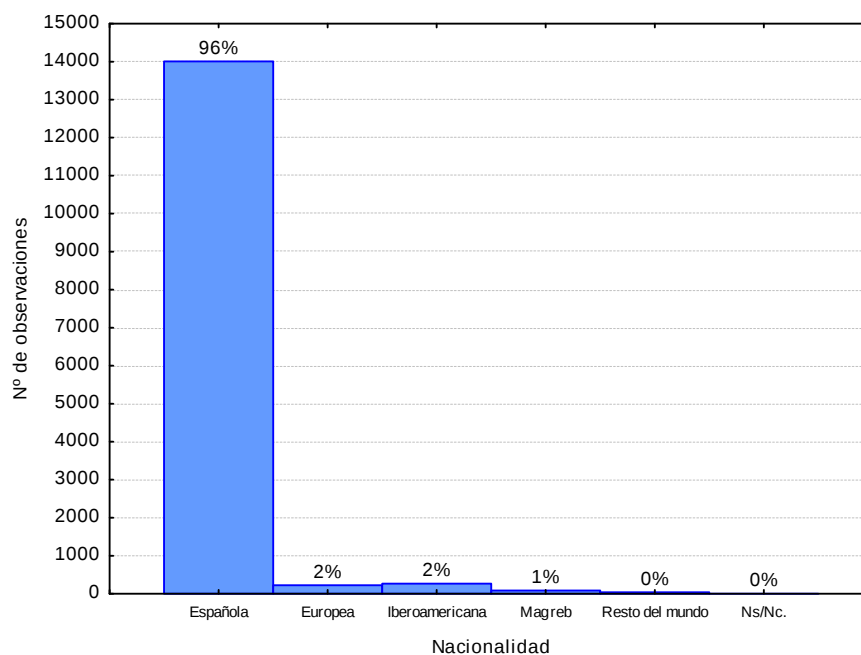
Figura 44. Frecuencias relativas del estado civil categorizado por sexo.



5. Nacionalidad

El 96% de nuestra muestra está compuesta por trabajadores españoles. El resto se reparte entre europeos, iberoamericanos y magrebíes. Pero la representación de los trabajadores extranjeros es muy baja (Figura 45). Por ello **la nacionalidad normalmente no será una variable relevante en nuestro estudio.**

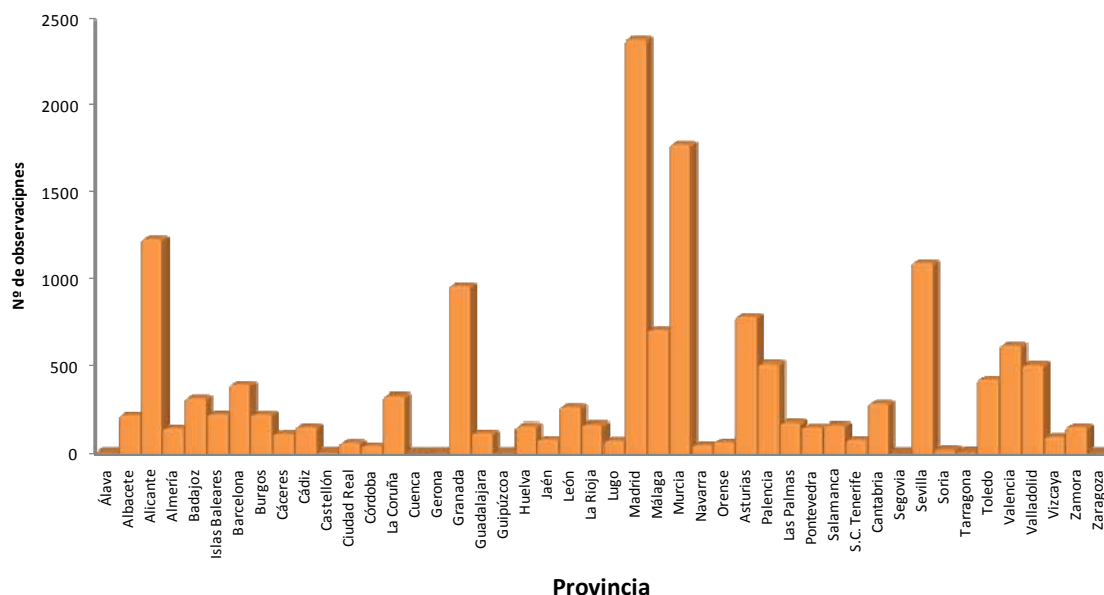
Figura 45. Histograma de la nacionalidad.



6. Provincia y Comunidad Autónoma donde recibe la baja

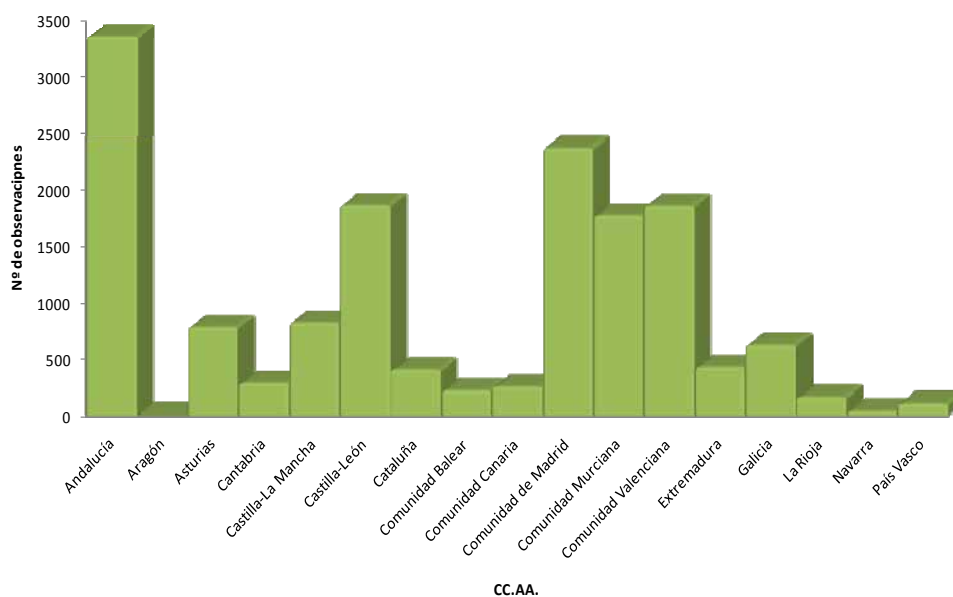
Podemos observar que nuestra muestra es muy heterogénea en lo que se refiere a la provincia en que recibió la baja el trabajador. En buena parte esto se debe a la presencia dispar que tenía Ibermutuamur por provincias en los años en que fue realizada la recogida de los datos. Otros factores como el tamaño de la población en la provincia, la implicación de los investigadores de campo que participaron (médicos de Contingencias Comunes), el tipo de ocupaciones y de patologías predominantes en cada provincia, etc. han podido tener, sin duda, alguna influencia en la recogida tan dispar de los datos por provincia, pero seguramente no tan importante como el primer factor comentado. Así, podemos observar en la Figura 46 que las dos provincias con más casos incluidos en el estudio son, precisamente, Madrid y Murcia, ambas con una presencia de Ibermutuamur históricamente importante. Le siguen provincias como Sevilla, Alicante, Granada y Asturias. Otras provincias que podemos destacar son: Valencia, Valladolid, Málaga, Palencia, Toledo o Barcelona. En otras como, por ejemplo, Castellón, Guipúzcoa, Segovia, Cuenca o Zaragoza, la recogida de datos ha sido muy escasa.

Figura 46. Histograma de casos por provincia.



Si lo miramos por Comunidades Autónomas (Figura 47), son las comunidades Andaluza, Madrileña, Valenciana, Murciana y Castellano-Leonesa las más representadas. Comunidades como la Aragonesa, Navarra y País Vasco apenas han aportado casos a nuestro estudio.

Figura 47. Histograma de casos por CC.AA.

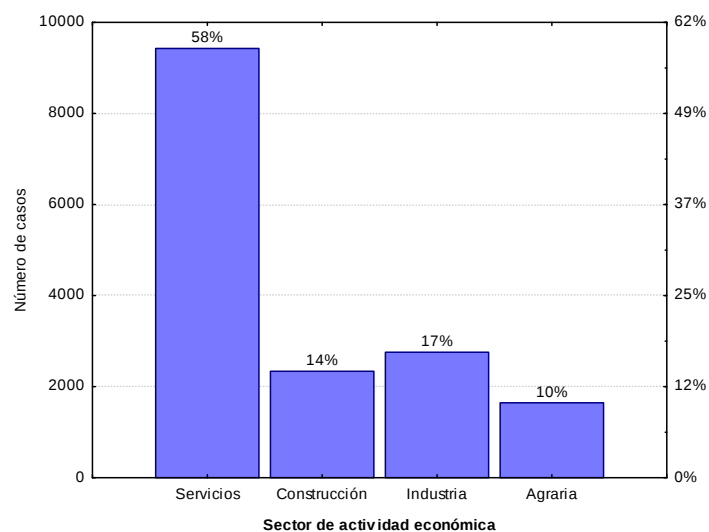


Variables relacionadas con los datos de tipo laboral

1. Sector de actividad económica de la empresa

Por sectores laborales, más de la mitad de los pacientes (58%) con baja por ITCC pertenecían al sector de Servicios (Figura 48). El resto se reparten entre los sectores de Industria (17%), Construcción (14%) y Agrario (10%).

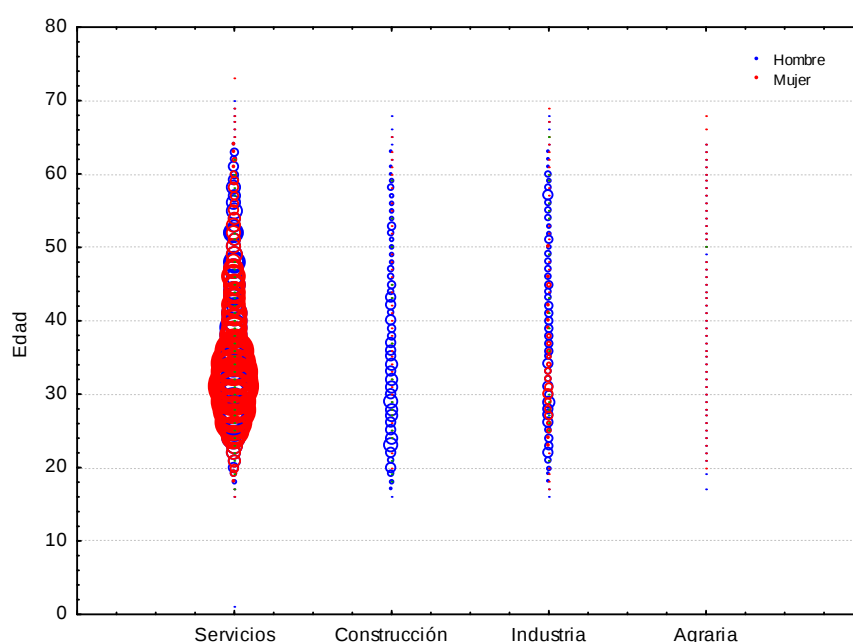
Figura 48. Frecuencias relativas por sector de actividad económica.



Sexo y edad por sector de ocupación

Como podemos observar en el diagrama de la Figura 49, la mayor parte de la muestra se encuentra en el sector de Servicios. En dicho sector hay tanto hombres como mujeres de todas las edades (del rango muestral). Sin embargo, es especialmente llamativa la presencia en dicho sector de mujeres (el 80,9% de todas las mujeres de la muestra), especialmente en las edades comprendidas entre los 25 y los 45 años; mientras que su presencia es mucho más escasa en el sector Industrial (13,3%) y casi desaparecen por completo en los sectores Agrario (2,6%) y de la Construcción (3,1%). Los hombres, si bien son más frecuentes en el sector de Servicios (50,9% sobre el total de hombres de la muestra), tienen también una presencia notable en los sectores de la Industria (21,9%) y de la Construcción (24,8%); tan solo el 2,5% restante trabajan en el sector agrario. Encontramos diferencias estadísticamente significativas ($F=26,74$; $p<0,01$) en la *edad* para los factores *rama* (sector de actividad económica) y *sexo*; y también una relación estadísticamente significativas ($\chi^2=1737$; $p<0,01$) entre las variables *rama* y *sexo*.

Figura 49. Número de observaciones por sector de actividad y por edad categorizados por sexo.



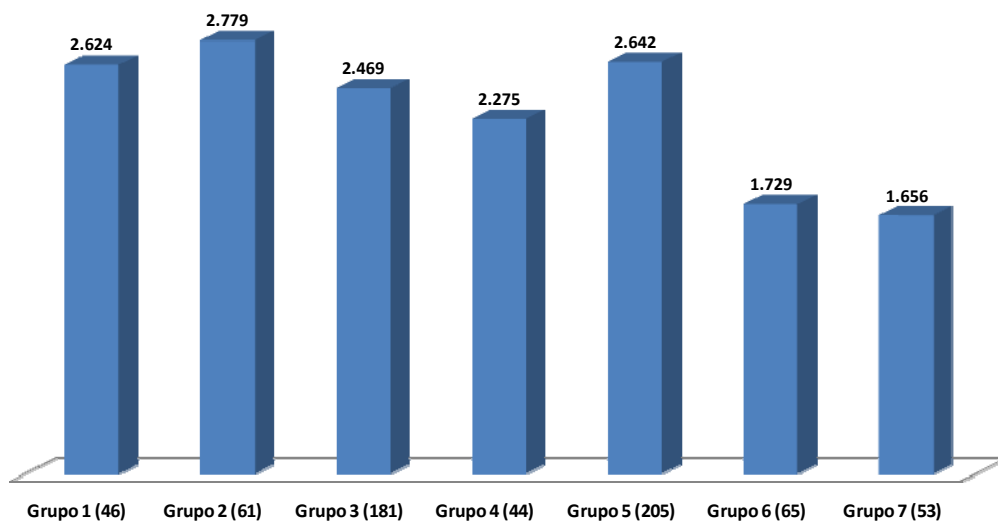
2. Actividad económica de la empresa

Como vimos en la página 66, la variable correspondiente a la actividad económica contienen tantas categorías diferentes (656) que para poder realizar los

análisis procedimos a agrupar dichas categorías en función de su efecto sobre la duración de la ITCC; de modo que categorías con efectos similares sobre la duración de la ITCC se clasifican en el mismo grupo y presentan efectos claramente diferenciados de las categorías clasificadas en los demás grupos. Hemos reducido, por lo tanto, todas las categorías a solo 7 grupos. Las categorías incluidas en cada uno de los grupos pueden consultarse en el Anexo 9.

En el diagrama de barras de la Figura 50 pueden apreciarse el número de observaciones incluidas en cada uno de dichos grupos y (en el eje X entre paréntesis) el número de categorías clasificadas en los mismos. Vemos que todos ellos contienen suficiente número de observaciones para realizar los análisis pertinentes.

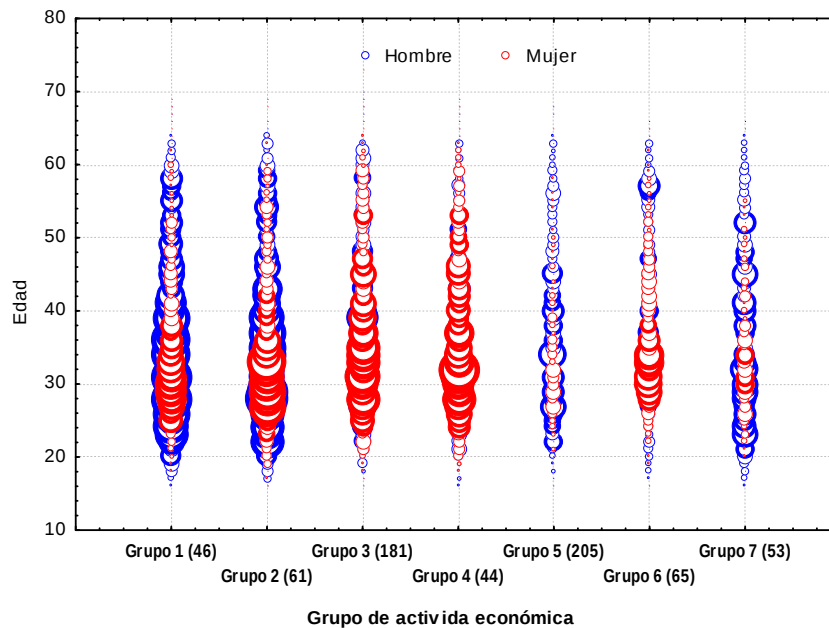
Figura 50. Número de observaciones por grupo de actividad económica.



Sexo y edad por Grupos de actividad económica de la empresa

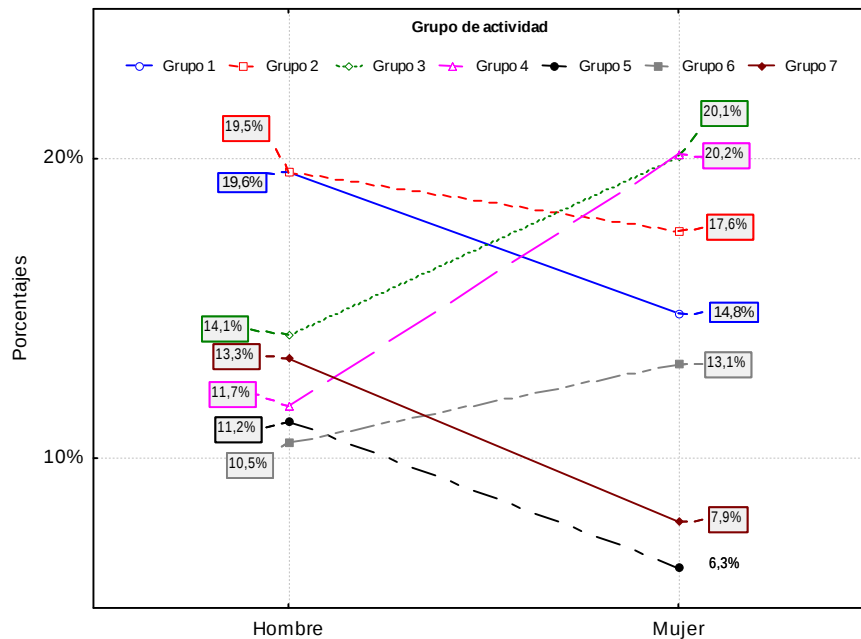
En la Figura 51 podemos observar las frecuencias absolutas para cada categoría de *sexo* y de *Actividad económica agrupada* y su relación con la edad del paciente. Vemos que en todos los grupos y rangos de edad están representados ambos sexos pero las frecuencias absolutas varían de unos a otros. Así, por ejemplo, en el grupo 4 predominan las mujeres, sobre todo las jóvenes y de edades medias, mientras que en el grupo 5 predominan los hombres de todas las edades. En todos los grupos están representadas todas las edades, aunque también varían las frecuencias según la edad y el grupo.

Figura 51. Número de observaciones por grupo de actividad y edad categorizadas por sexo.



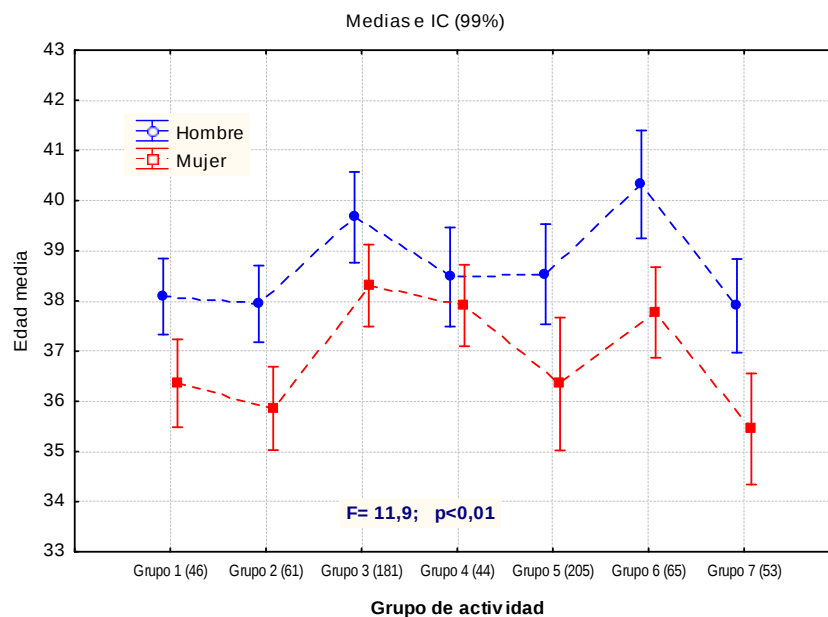
Como en la muestra tenemos más hombres (58%) que mujeres (42%), según vimos anteriormente (ver “Edad y sexo de los pacientes incluidos en el estudio” en la página 18), es preciso comparar las frecuencias relativas para cada sexo. En la Figura 52 podemos ver representada las frecuencias relativas (con relación al total de cada categoría de *sexo* y, por lo tanto, directamente comparables) de cada grupo de *actividad económica* para cada categoría de *sexo*. Como podemos apreciar, ambos sexos están presentes en todas las categorías pero en proporciones variables. Y, según el grupo, predomina uno u otro sexo, a saber: las mujeres en los grupos 3, 4 y 6; los hombres en los grupos 1, 2, 5 y 7.

Figura 52. Frecuencias relativas comparadas para hombre y mujer para cada grupo de actividad.



Al comparar las medias de edad para cada grupo de actividad y cada uno de los géneros, observamos diferencias estadísticamente significativas ($F=11,9$; $p<0,01$) entre la mayor parte de ellos (los grupos y los sexos), aunque no para todos (Figura 53).

Figura 53. Edades medias por sexo y grupo de actividad económica.

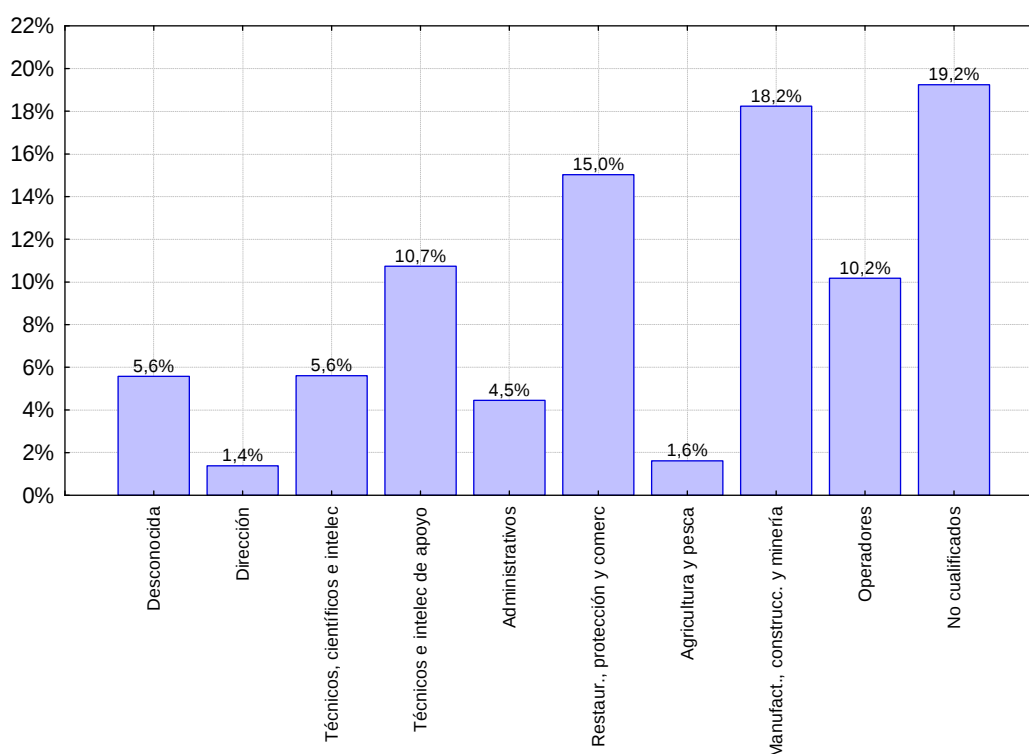


Como conclusión, encontramos diferencias en las distribuciones para *edad* y *sexo* ente los diferentes grupos, si bien hay una representación de todas las edades y de ambos géneros en todos ellos.

3. Ocupación

Por grupos de ocupación encontramos que buena parte de los pacientes son “no cualificados”, “operadores” o trabajan en manufacturas, construcción y minería (el 47,6% de la muestra se clasifica en alguna de estas categorías; Figura 54). También son relativamente frecuentes los trabajadores de “Restauración, protección y comercio” (15%) y “Técnicos e intelectuales de apoyo” (10,7%). Puestos de dirección (1,4%) así como la agricultura y pesca (1,6%) son poco frecuentes en nuestra muestra.

Figura 54. Frecuencias relativas para la ocupación (según CNO93)

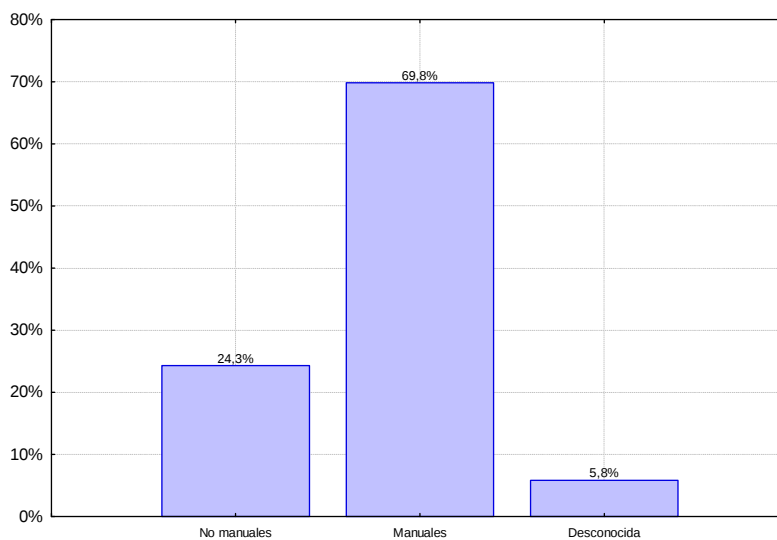


De acuerdo con los datos de nuestra muestra, tan solo el 1,8% de los trabajadores de la muestra están pluriempleados (ver pág. 94).

Trabajadores manuales (o de “cuello azul”) y no manuales (o de “cuello blanco”)

Como podemos intuir observando la Figura 55, casi el 70% de la muestra son trabajadores con tareas que requieren un mayor esfuerzo físico y que denominaremos como “de tipo manual” (o “de cuello azul”). Los trabajadores que denominaremos “de tipo no manual”, o también “de cuello blanco” (aquellos que suelen realizar tareas que requieren generalmente un mayor esfuerzo de tipo intelectual) son mucho menos frecuentes (24,3%) en la misma.

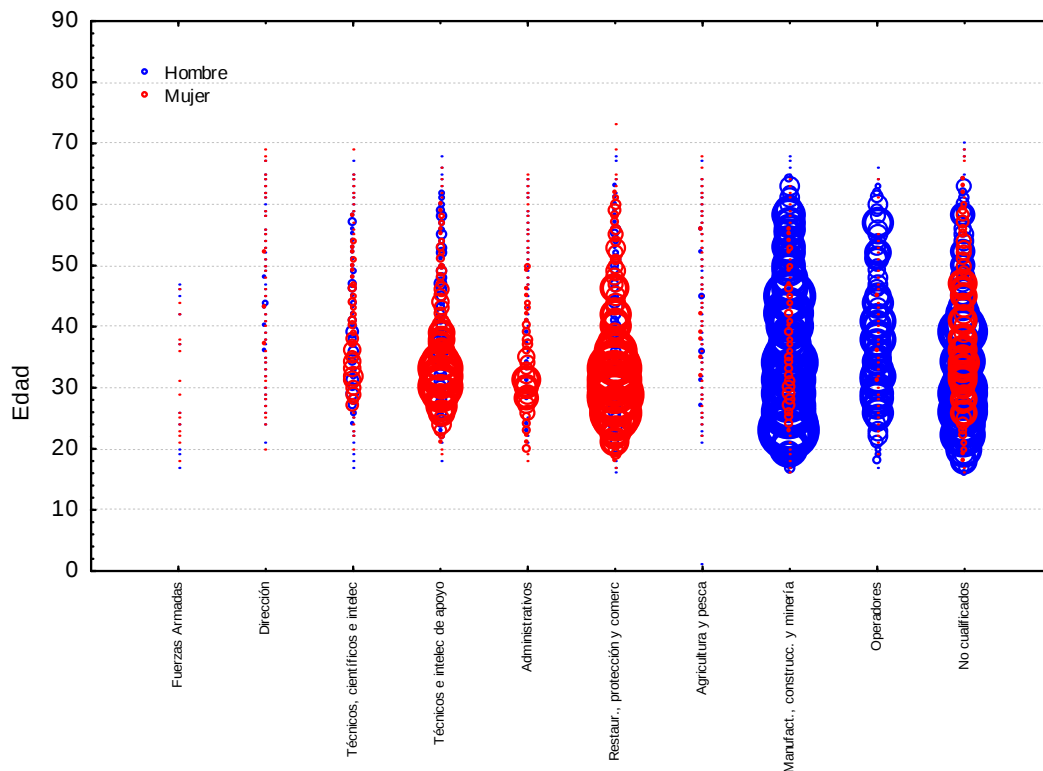
Figura 55. Frecuencias relativas para trabajadores “manuales” y “no manuales”.



Ocupación por edades y sexo

En la Figura 56 podemos observar cómo existe una relación entre el sexo, la edad y la ocupación entre los trabajadores de nuestra muestra. Así, las mujeres entre los 20 y los 40 años suelen tener ocupaciones agrupadas bajo las categorías de “Restauración, protección y comercio”, de “Técnicos e intelectuales de apoyo” o de “Administrativos”; mientras que los hombres de diferentes edades se encuentran más en ocupaciones clasificadas bajo las categorías “Manufacturas, construcción y minería”, “Operadores” y “No cualificados”. También encontramos buen número de mujeres de todas las edades en esta última categoría.

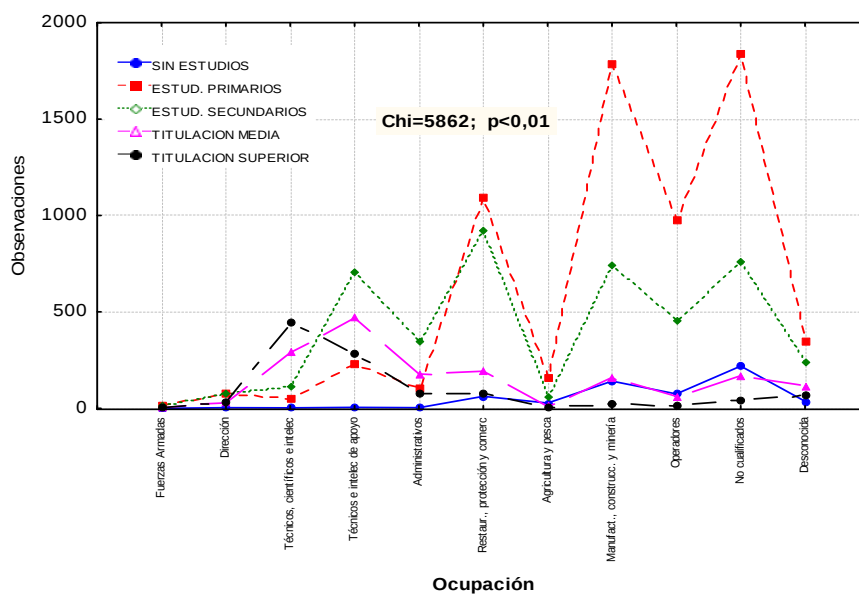
Figura 56. Distribución de los casos por ocupación, sexo y edad.



Ocupación y nivel de estudios

Como podríamos intuir existe una relación estadísticamente significativa ($\chi^2=5862$; $p<0,01$) entre la ocupación y el nivel de estudios (Figura 57). Y, por lo tanto, como hemos visto anteriormente, entre el sexo y la edad y estos dos.

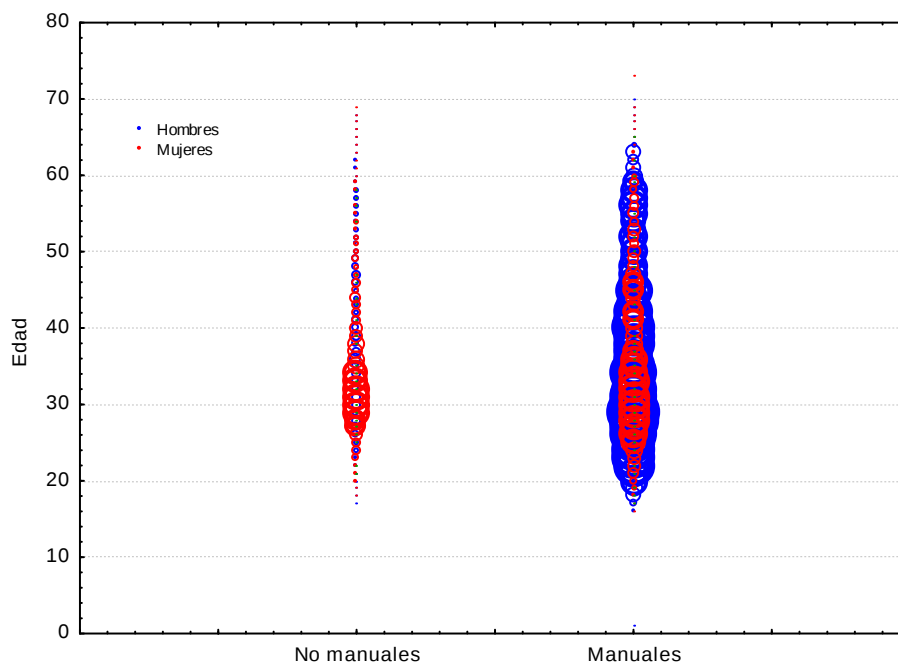
Figura 57. Frecuencias absolutas comparadas para ocupación y nivel de estudios.



Trabajadores “manuales” y “no manuales” y sexo

Como ya hemos visto, en nuestra muestra tenemos más hombres (58%) que mujeres (42%) y más trabajadores manuales (69,8%) que no manuales (24,3%). Al tener en cuenta la interacción entre ambas variables, encontramos que la proporción de trabajadores manuales es mayor entre los hombres (83,2%) que entre las mujeres (61,6%); mientras la proporción de los no manuales es superior entre las mujeres (38,4%) comparadas con los hombres (16,8%). Por lo tanto, en nuestra muestra hay una mayor tendencia entre las mujeres hacia los trabajos no manuales y entre los hombres hacia los manuales (Figura 58).

Figura 58. Distribución de los trabajadores “manuales” y “no manuales” por sexo y edad.



4. Tipo de contrato

El tipo de contrato más habitual entre los trabajadores de nuestra muestra es el indefinido (61%). Hay otro 38% de trabajadores con contrato temporal. Y un 1% en formación. Ver la Figura 59. En la Figura 60, de la página 87, podemos observar las distribuciones de frecuencias absolutas para la *edad* y el *tipo de contrato* categorizadas por *sexo*.

Figura 59. Frecuencias relativas para el tipo de contrato.

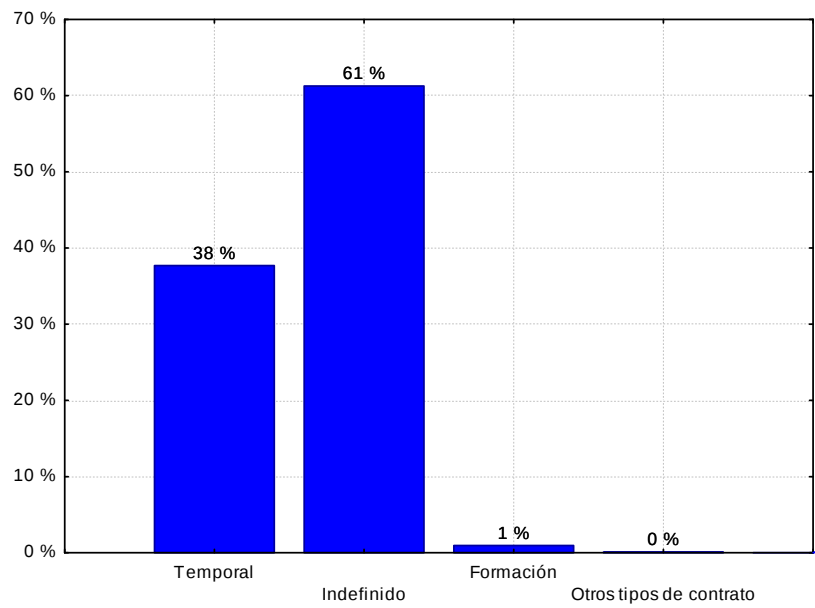
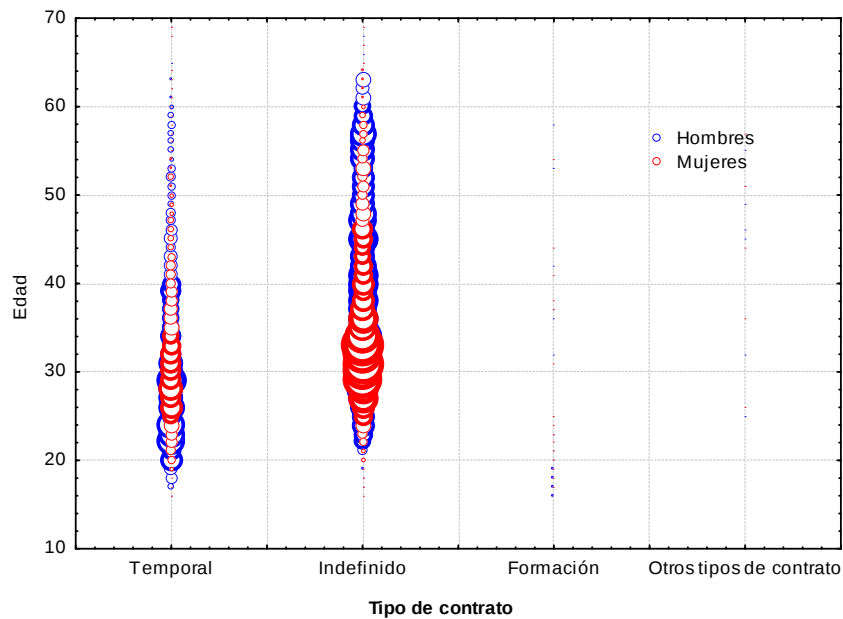


Figura 60. Frecuencias absolutas por edad, sexo y tipo de contrato.

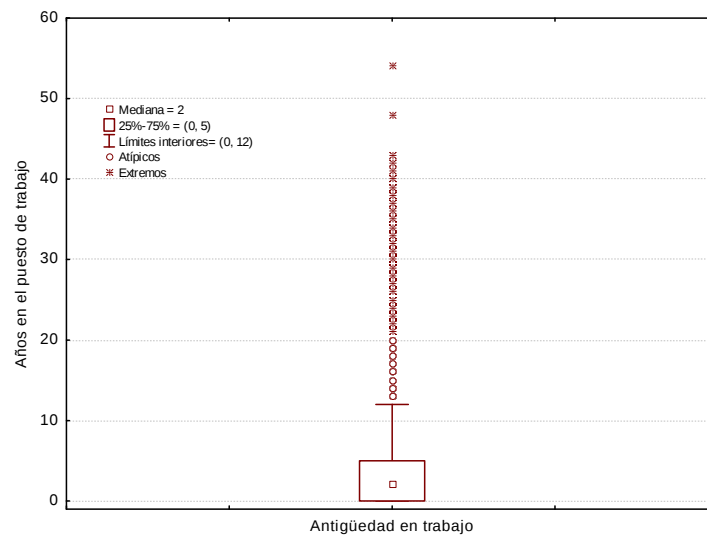


Se aprecia cierta tendencia, en los contratos temporales, hacia edades más jóvenes, si bien los encontramos a cualquier edad.

5. Antigüedad en el puesto de trabajo

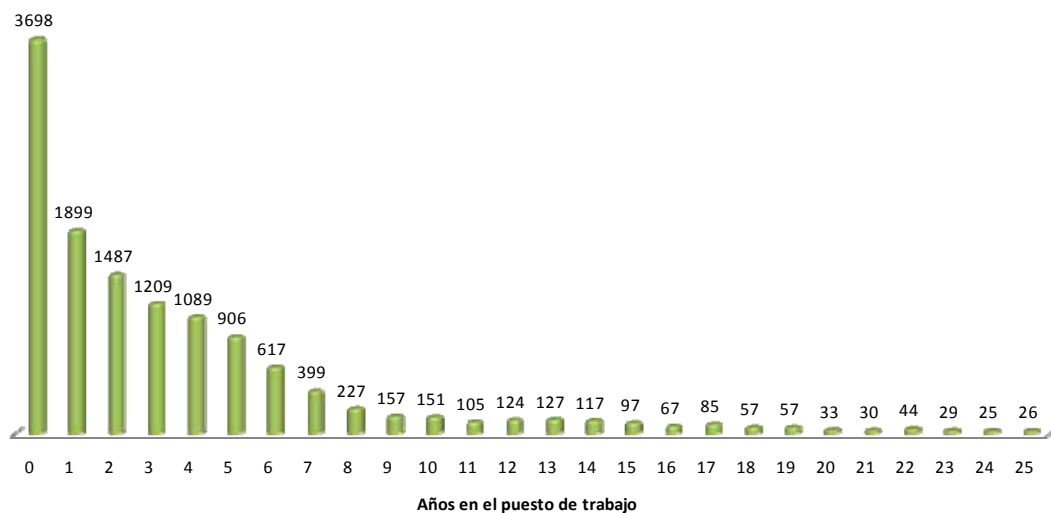
La mayor parte de los trabajadores de nuestra muestra han permanecido pocos años en su empresa. La mediana muestral son 2 años y los cuartiles son 0 (P_{25}) y 5 años (P_{75}). Hay casos con permanencias muy largas en la empresa, por encima de los 12 años, pero en nuestra muestra pueden considerarse como valores atípicos o extremos, aunque se trate de valores correctos (Figura 61).

Figura 61. Distribución de los casos por años de antigüedad en la empresa.



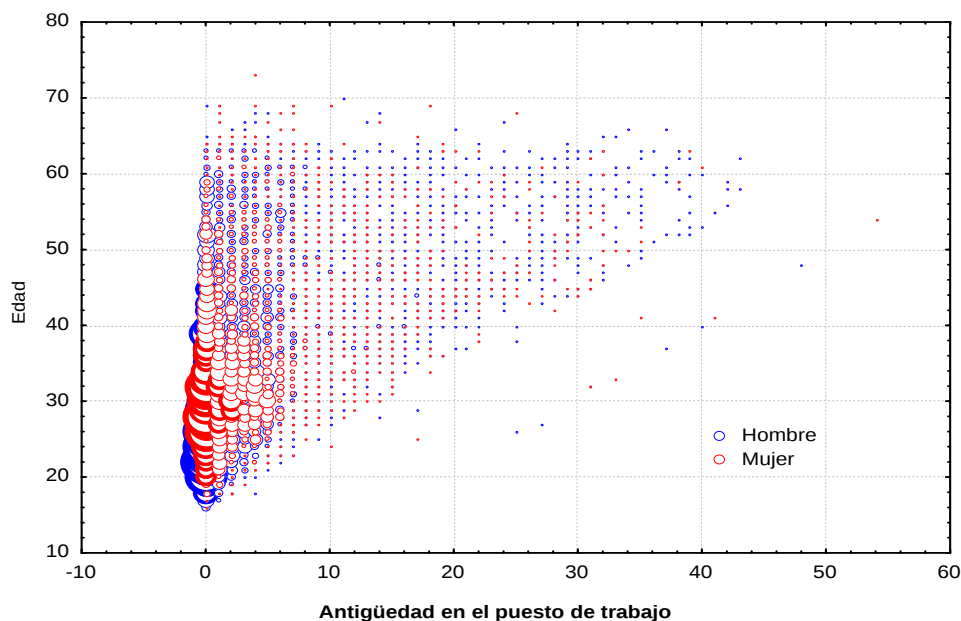
Observando el histograma de la Figura 62 vemos que 3.698 casos llevan menos de 1 año en la empresa (28% del total). Luego descenden las frecuencias rápidamente entre 1 año y 9 años y a partir de los 10 descenden de forma más moderada.

Figura 62. Histograma por años de permanencia en la empresa.



Una gran parte de los trabajadores llevan pocos años en la empresa. Así, el 78,4% ha estado 5 años o menos en la empresa actual. Esto podemos apreciarlo mejor en el diagrama de la Figura 63. En el mismo, podemos observar las distribuciones por edad, antigüedad en la empresa y sexo del paciente. Se puede apreciar claramente cómo un número muy importante de los casos se acumulan en los primeros años de antigüedad tanto para los hombres como para las mujeres y para todas las edades.

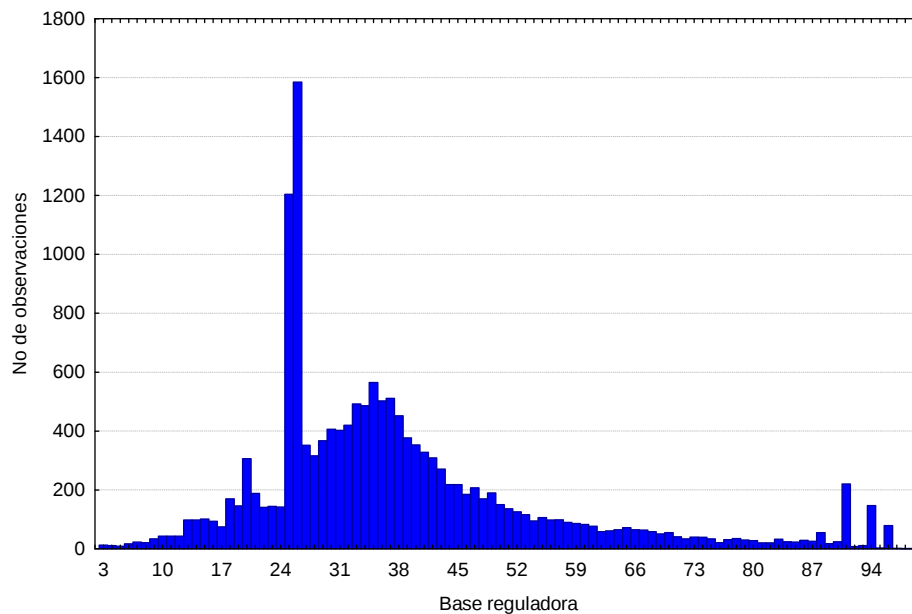
Figura 63. Antigüedad en la empresa, por edad y sexo.



6. Base reguladora

La base reguladora sigue una distribución que se muestra en el histograma de la Figura 64. Aunque se distribuyen a lo largo de todo el rango de valores entre un mínimo de 3 y un máximo 119, observamos tres niveles que llaman la atención por la cantidad de casos que se acumulan en torno a los mismos. El más frecuente coincide con valores en torno a la moda que es de 26€. El siguiente nivel está en torno a los 35€. Otro nivel interesante estará próximo a los 90€.

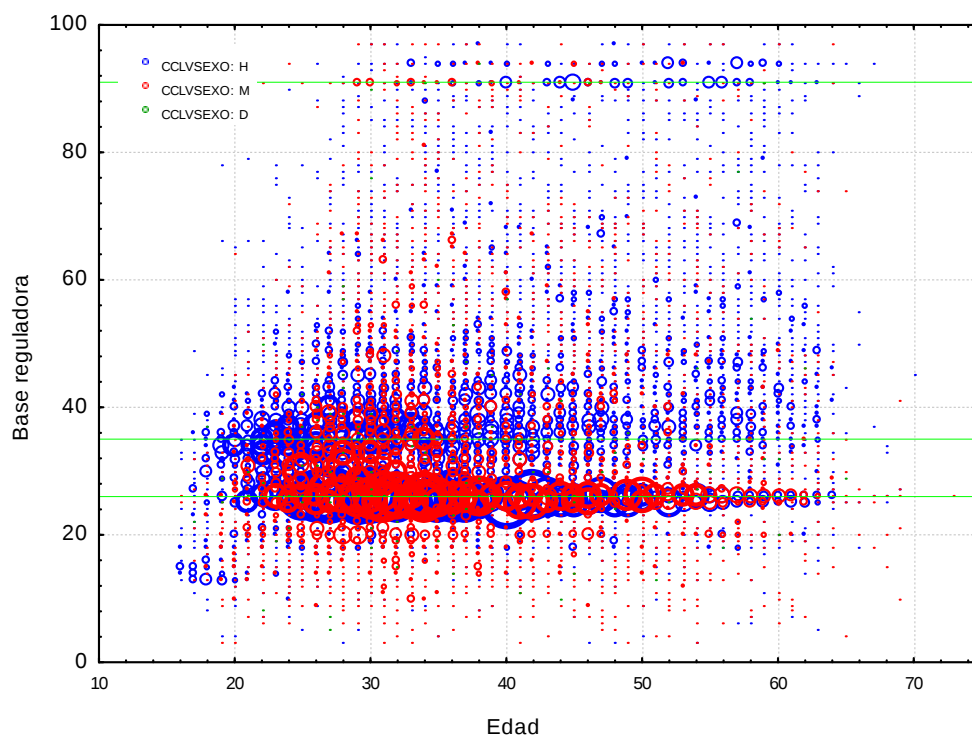
Figura 64. Histograma de la base reguladora.



Base reguladora por edad y sexo

Como podemos observar en la Figura 65, los valores tienden a acumularse preferentemente en torno a los 3 niveles que hemos comentado anteriormente. En torno al primero, el de los 26€, encontramos buena parte de los casos de ambos sexos. Las mujeres parecen situarse sobre todo en las proximidades de este nivel. De hecho, las proporciones de mujeres en valores próximo al mismo son superiores que para los hombres. Otro grupo grande de casos estaría en valores próximos al segundo nivel, el de los 35€. Hay menos mujeres en este nivel. Las proporciones son mayores para los hombres. Muy lejos estaría el tercer nivel, el de los 91€, y, además encontramos muy pocos casos en este nivel. Aunque hay algunas mujeres en este nivel, las proporciones también son menores para las mismas. Consultar Anexo 12 (Distribución de la base reguladora por sexo) en la página 253.

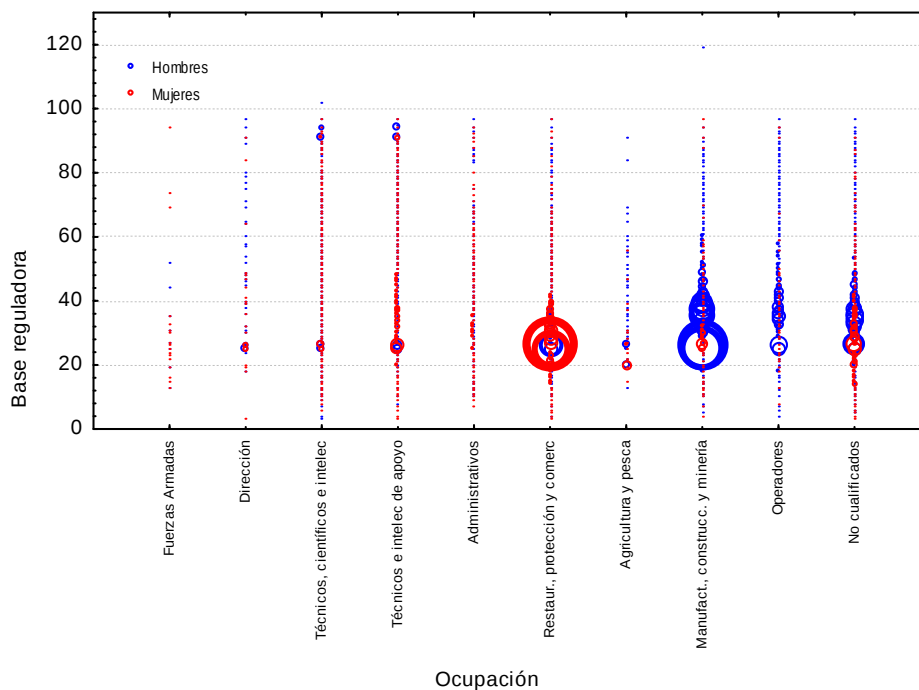
Figura 65. Distribución de los valores de la base reguladora por edad y categorizados por el sexo.



Base reguladora por ocupación y sexo

En la Figura 66 podemos apreciar que las mujeres, si bien se encuentran distribuidas por todas las ocupaciones y por todo el rango de la base reguladora, se concentran especialmente en valores próximos a la moda, 26€, y en la categoría de “Restauración, protección y comercio”. Los hombres, en cambio, se concentran especialmente en “Manufacturas, construcción y minería” y en los niveles cercanos a 26€ y 35€. En torno al nivel más alto, de 91€, hay tanto hombres como mujeres aunque en mayor proporción los primeros. Además gran parte de los casos se concentran en dos categorías: “Técnicos, científicos e intelectuales” y “Técnicos e intelectuales de apoyo”.

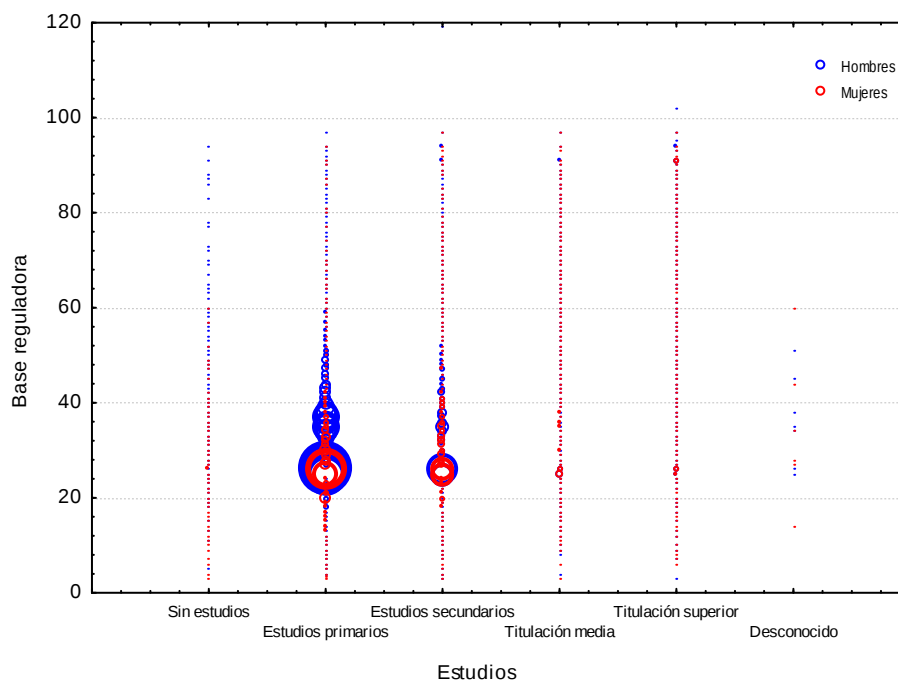
Figura 66. Distribución de la base reguladora por ocupación y sexo.



Base reguladora por nivel de estudios y sexo

La base reguladora también está relacionada con el nivel de estudios como veremos a continuación. En la Figura 67 puede apreciarse que tanto las mujeres como los hombres se concentran mayormente en niveles de base reguladora bajos (26€ y 35€) y se corresponden con estudios primarios y secundarios. Abundan más las mujeres con titulaciones medias y superiores que los hombres con dichas titulaciones, como ya vimos anteriormente y éstas se distribuyen por todo el rango de valores de la base reguladora.

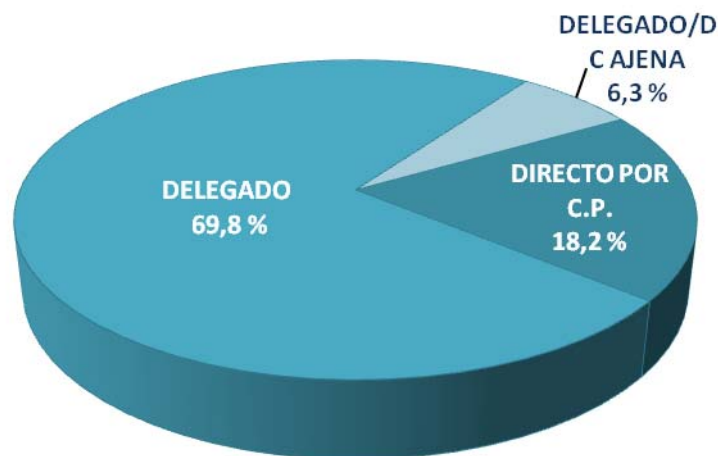
Figura 67. Frecuencias absolutas de observaciones de la base reguladora por nivel de estudios y sexo



7. Modo de pago

El 69,8% de los pacientes de nuestra muestra estaban en pago “Delegado”, es decir, el trabajador está en baja por ITCC y continúa vinculado contractualmente a la empresa para la que trabaja. Otro 18,2% está en modo de pago “Directo por cuenta propia”. Esta categoría se corresponde con aquellos trabajadores que son autónomos o pertenecen a los regímenes especiales Agrario o de Trabajadores del Mar. Los restantes, el 6,3%, pertenecen al modo de pago “Delegado/directo por cuenta ajena”. Éstos últimos son trabajadores que comienzan su periodo de baja con modo de pago “Delegado” pero durante la misma se produce la extinción de su contrato laboral con la empresa por cualquier causa (finalización del contrato, despido, etc.) o bien la empresa no puede hacer frente al pago de su IT (Figura 68).

Figura 68. Modos de pago.



No encontramos diferencias estadísticamente significativas en cuanto a los modos de pago cuando lo analizamos por *sexo*; si las encontramos ($p < 0,01$) al analizarlo teniendo en cuenta otras variables como son: *edad, número de hijos, nivel de estudios, antigüedad en la empresa, tipo de contrato, sector de actividad económica, ocupación, trabajador manual/no manual y provincia*.

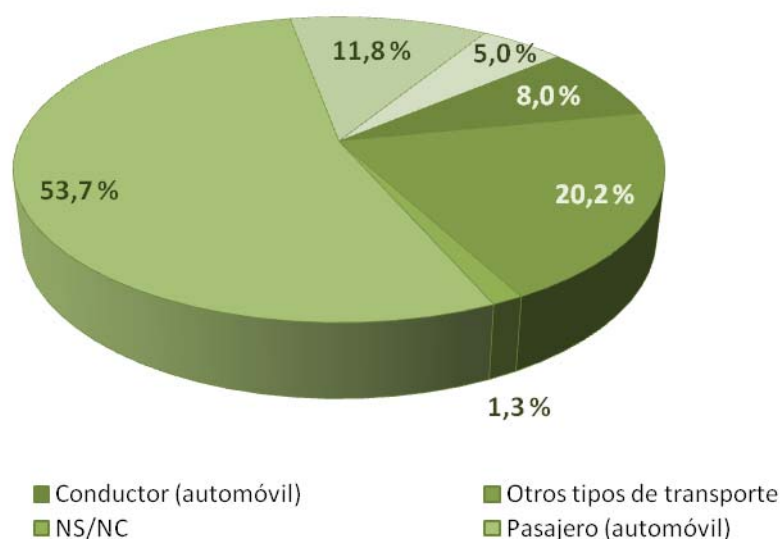
8. Pluriempleo

Solo estaban pluriempleados en el momento de recoger los datos del estudio 286 trabajadores; esta cifra representa menos del 1,8% de toda la muestra.

9. Tipo de transporte en el que se dirige al trabajo

La mayoría de los trabajadores de nuestra muestra se desplaza diariamente al trabajo preferentemente en automóvil en calidad de pasajero (53,7%); en transporte público lo hace el 11,8% y otro 8% lo hacen conduciendo un automóvil. El 20,2% de los pacientes se trasladan utilizando otros medios de transporte. Ver la Figura 69.

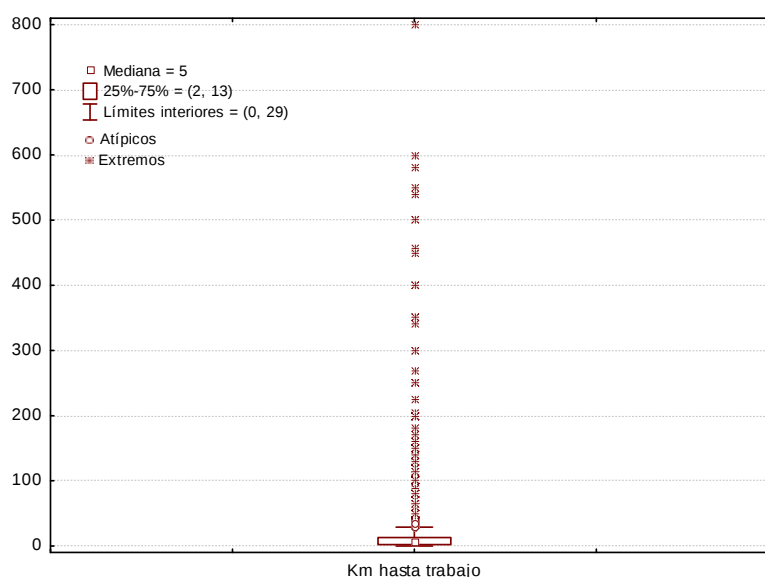
Figura 69. Medios de transporte hasta el puesto de trabajo.



10. Distancia hasta el centro de trabajo

La mayoría de los trabajadores de la muestra viven relativamente cerca del puesto de trabajo. Así el 25% de los mismos recorren una distancia igual o inferior a 2 Km; el 50% tienen que desplazarse 5 km o menos; el 75% hasta 13 km. Algunos recorren distancias mucho más largas, mayores de 30 Km; se trata de casos atípicos y extremos y raros. En el siguiente diagrama (Figura 70) podemos apreciar la distribución, la mediana y los cuartiles de la distancia recorrida hasta el centro de trabajo.

Figura 70. Distancia hasta el centro de trabajo.

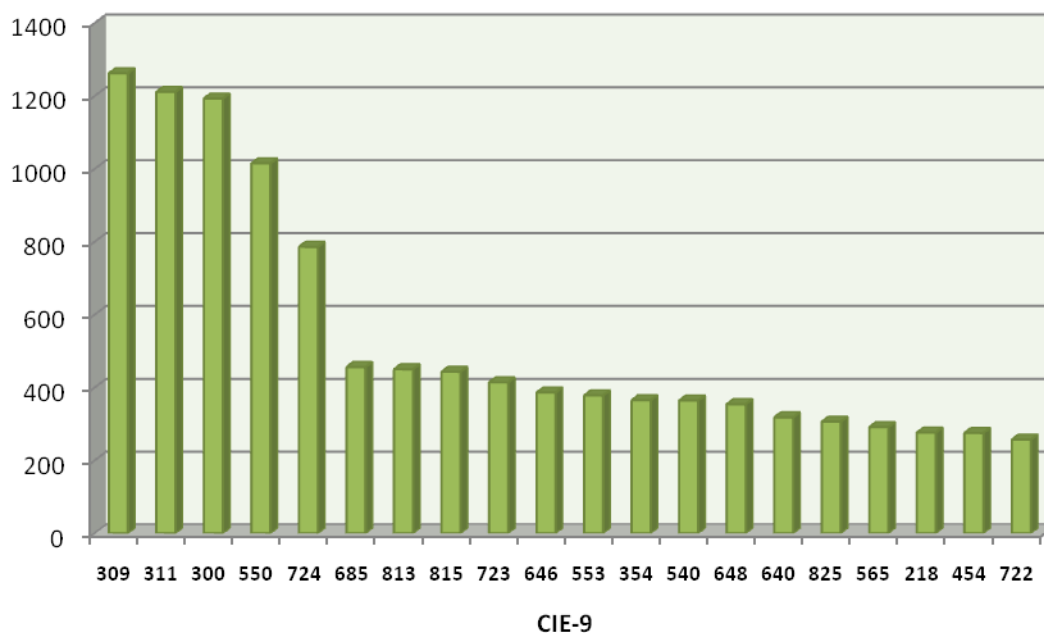


Variables relacionadas con los datos de tipo clínico

1. Diagnósticos (según CIE-9)

En la muestra que hemos recogido encontramos 677 categorías diferentes para la variable *diagnóstico*. No es fácil analizar y, menos aún, interpretar un número tan grande de categorías. Para simplificar, en el presente documento hemos reducido los diagnósticos a los que se corresponden con las categorías (códigos de tres dígitos) del CIE-9. Aún así encontramos 122 categorías diferentes que pueden consultarse en la tabla correspondiente del Anexo 11, junto con sus prevalencias (en nuestra muestra). En la Figura 71 podemos ver las 20 primeras categorías ordenadas según la frecuencia de observaciones encontradas.

Figura 71. Número de observaciones para cada diagnóstico (según CIE-9).

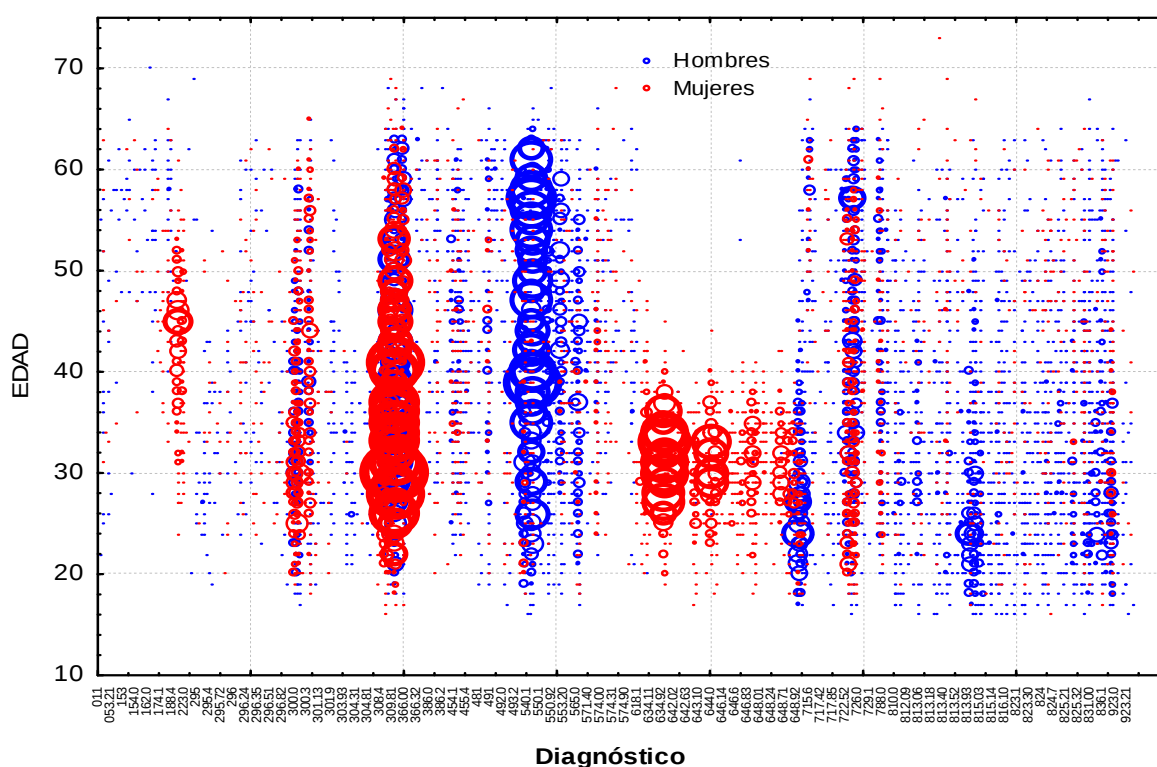


Frecuencias de los diagnósticos por sexo y edad

Como cabría esperar intuitivamente, las distribuciones por edad y sexo varían con relación a los diferentes diagnósticos. Este hecho es especialmente intenso para algunos diagnósticos. Así, por ejemplo, el diagnóstico 640 (“Hemorragias en fase temprana del embarazo” según CIE-9) solo ocurre en mujeres en edad fértil. Las enfermedades mentales como los diagnósticos 300, 309 y 311 son más frecuentes en las mujeres que en los hombres de nuestra muestra y aparecen en todas las edades.

En cambio el diagnóstico 550 ("Hernia Inguinal") casi solo se ha producido en hombres (el 95% de los pacientes con Hernia Inguinal han sido varones). Las diferencias que observamos pueden tener diversas causas. Así, el diagnóstico 640, por su propia naturaleza, solo puede producirse en mujeres; en cambio, el 550 puede tener relación con el tipo de tareas realizadas en el trabajo o con otras actividades que realicen con mayor frecuencia los hombres; las enfermedades mentales pueden relacionarse con diversos factores como el estrés laboral, entorno familiar, situación socio-económica, etc. Algunos diagnósticos aparecen con frecuencias similares para ambos sexos. Algo parecido sucede con relación a la edad. La naturaleza del propio proceso o de los factores que lo causan puede condicionar el hecho de que se produzca preferentemente en determinados rangos de edad o bien a cualquier edad. En el gráfico siguiente pueden apreciarse visualmente dichas diferencias (Figura 72).

Figura 72. Número de observaciones por diagnóstico, edad y sexo.



2. Comorbilidad

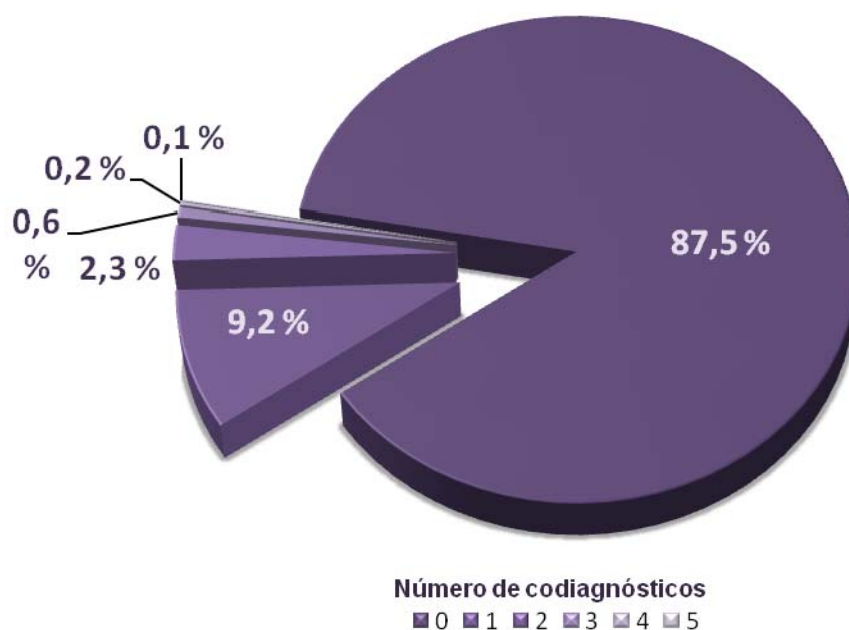
La presencia de otros trastornos patológicos, coincidiendo en el tiempo con el diagnóstico principal, puede alargar la duración del proceso y, por lo tanto, de la baja por ITCC. Aunque en nuestra muestra no son muy frecuentes (12,5%), si son

importantes, como veremos más adelante, por su influencia sobre la duración de la ITCC. La gran mayoría de los pacientes (9,2% sobre el total de la muestra) que presentaron comorbilidad, presentaron solo un codiagnóstico (ver Tabla 6 y Figura 73). Otro 2,3% de la muestra tenían 2 codiagnósticos. Solo el 0,9% tenían 3 o más.

Tabla 6. Comorbilidad.

Nº Codiagnósticos	Observaciones	%
0	14.158	87,5 %
1	1.487	9,2 %
2	376	2,3 %
3	102	0,6 %
4	36	0,2 %
5	15	0,1 %

Figura 73. Comorbilidad.



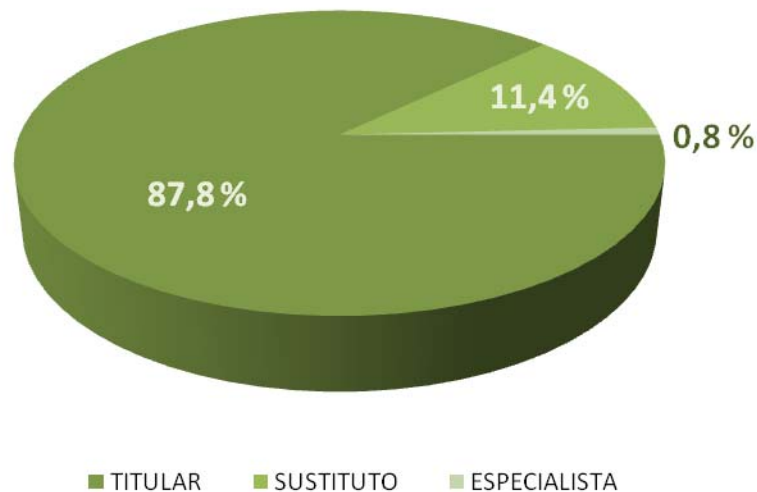
3. ¿Qué médico le dio la baja laboral por ITCC?

En el 87,8% de los casos la baja fue extendida por su médico titular. Solo un 11,4% la recibió de un médico sustituto y el 0,8% restante de un especialista (Tabla 7 y Figura 74).

Tabla 7. Médico que le dio la baja.

	Observaciones	%
TITULAR	12.882	87,8 %
SUSTITUTO	1.668	11,4 %
ESPECIALISTA	115	0,8 %

Figura 74. Médico que le dio la baja.



4. Entidad responsable del tratamiento

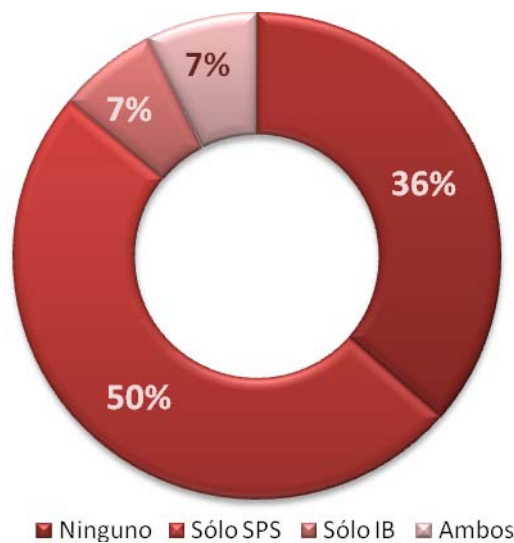
En el protocolo del estudio se creó un listado de tratamientos que se iban a recoger y estudiar con más detalle. Entre estos tratamientos estaban todos los que son de tipo quirúrgico, rehabilitación, psicoterapia y otras intervenciones. Quedaron fuera de dicho listado los tratamientos farmacológicos (exceptuando la quimioterapia, los psicofármacos y los agentes esclerosantes) y otros más conservadores como los relacionados con el estilo de vida (dieta, ejercicio, reposo, etc.). Estos tratamientos que se excluyeron del estudio se agrupan en esta variable bajo la categoría "Ninguno". Para los tratamientos incluidos en el estudio se recogió quién fue la entidad responsable de aplicarlo: el Servicio Público de Salud (SPS), Ibermutuamur o ambos.

El 36,4% de los procesos, de acuerdo con nuestros datos, recibieron otros tratamientos diferentes de los recogidos en el estudio; es decir, se agruparon bajo la categoría "Ninguno" de esta variable. Un 49,6% fue tratado por el Servicio Público de Salud y un 6,6% por Ibermutuamur. Finalmente, el 7,3% fue tratado por ambos (Tabla 8 y Figura 75).

Tabla 8. Frecuencias de pacientes derivados.

	Observaciones	%
Ninguno	5.895	36,4 %
Solo SPS	8.023	49,6 %
Solo IB	1.068	6,6 %
Ambos	1.188	7,3 %

Figura 75. Frecuencias de pacientes derivados.



Los tratamientos incluidos suelen ser los menos conservadores: tratamientos agresivos como los quirúrgicos, que requieren la intervención de un especialista como la psicoterapia o la rehabilitación. Esta variable puede estar relacionada, indirectamente, con: la gravedad del proceso o con dificultades inherentes al mismo, con procesos que requieren un tratamiento específico o conocimientos más especializados o medios más avanzados; también podría relacionarse con procesos con evolución tórpida o inesperada; a una mala respuesta al tratamiento; complicaciones sobrevenidas; comorbilidad; u otros factores. Todos ellos pueden influir sobre la duración del proceso mismo y de la baja asociada, además de añadir los periodos de tiempo necesarios para realizar las gestiones administrativas, desplazamientos, días en listas de espera para una prueba diagnóstica o para recibir un tratamiento, etc. Por ello, esta variable puede ser interesante como marcador relacionado con una serie de factores, en ocasiones desconocidos por el investigador, que pueden influir en la duración de la baja, y, con frecuencia, alargándola. Y, como acabamos de ver, estos tratamientos se

han aplicado en el 63,6% de los casos; con bastante frecuencia, por lo tanto. Más adelante veremos que **esta variable es relevante como factor que influye en la duración de la baja por ITCC.**

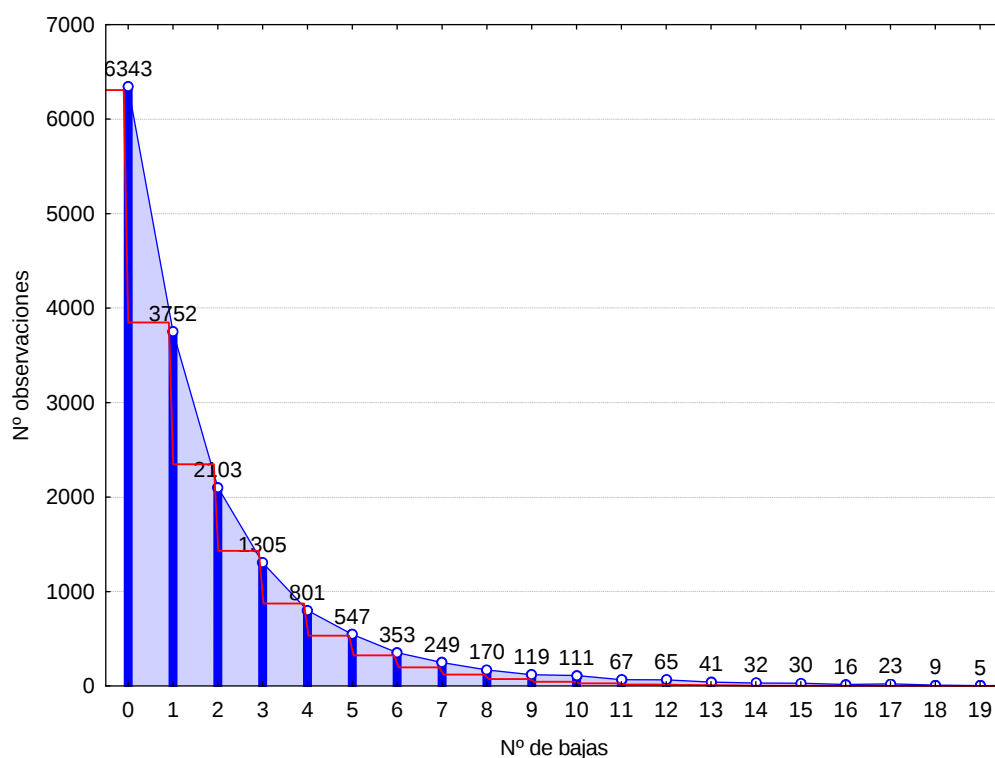
La Incapacidad Temporal por Contingencia Común (ITCC)

Bajas previas

1. Número de bajas previas

El *número de bajas previas* por cada paciente es una variable aleatoria que sigue una **distribución geométrica** que puede representarse por la siguiente fórmula: $f(x) = p \times (1 - p)^x$. Para nuestra muestra toma los siguientes valores: $p = 0,39$; $x = 0, 1, 2, \dots$. Si bien, utilizando pruebas de bondad de ajuste (Kolmogorov-Smirnov; Chi-cuadrado) los resultados ($p < 0,01$) no nos permiten afirmar que los datos de la muestra se ajusten a dicha distribución, es evidente, sin embargo, que sí se aproxima mucho a la misma, como podemos observar en la Figura 76. En la misma veremos de color azul la distribución de los valores de nuestra muestra y en la línea roja los valores que toma la distribución geométrica al calcularla con los valores que hemos fijado previamente para p y x . Podemos observar, también, cómo la frecuencia de las observaciones cae de forma geométrica a la vez que aumenta el número de las bajas previas.

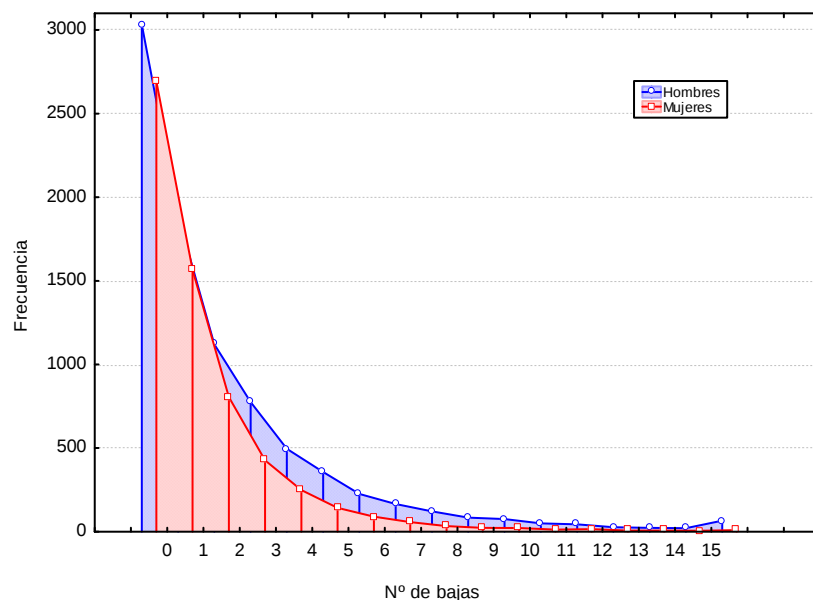
Figura 76. Distribución del número de bajas previas.



Número de bajas por sexo

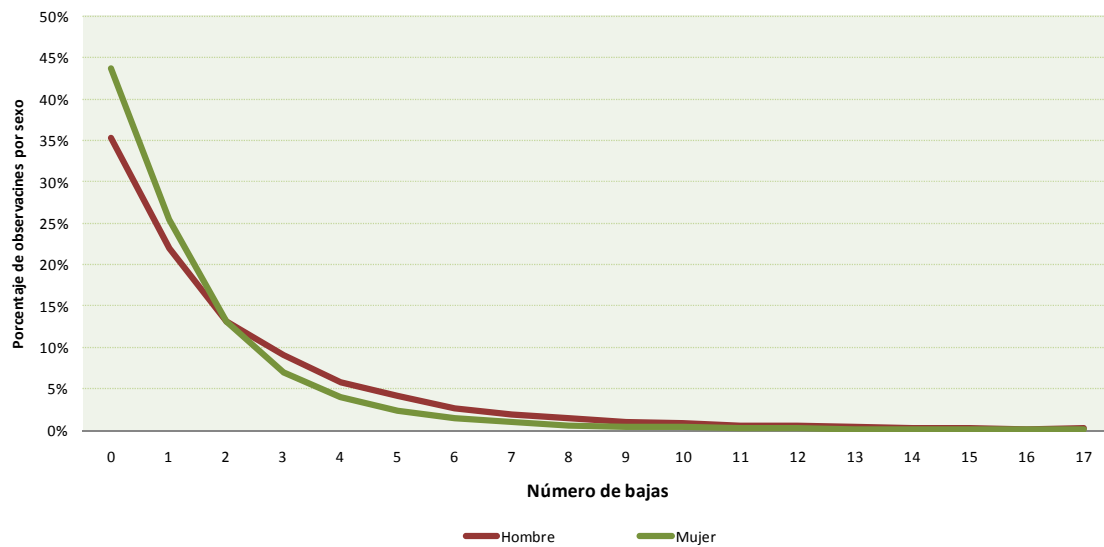
Al observar la misma distribución categorizada por *sexo* (Figura 77), podemos apreciar que, para todos los valores que toma la variable *número de bajas*, hay una menor frecuencia de casos para las mujeres, lo que puede explicarse, al menos en parte, por la menor proporción de éstas en la muestra. Pero también podemos apreciar que la curva cae de forma más pronunciada, lo que sugiere que para las mujeres el número de bajas previas tiende a ser menor que para los hombres. En el gráfico hay una pequeña separación entre ambas curvas para facilitar la comparación; en realidad, los puntos han de superponerse para ambas curvas.

Figura 77. Distribuciones por sexo del número de bajas previas.



Podemos apreciar mejor la mencionada diferencia en la Figura 78. En ella representamos las frecuencias relativas de las observaciones con relación al total de casos para cada sexo. Podemos ver cómo hay mayor proporción de mujeres con 0 bajas previas y también con 1. Pero las proporciones se igualan para 2 bajas y es mayor en los hombres a partir de 3. Para valores por encima de 10 vuelven a igualarse, debido a que se trata de casos muy poco frecuentes. Las diferencias observadas son estadísticamente significativas (K-S y χ^2 ; $p < 0,01$).

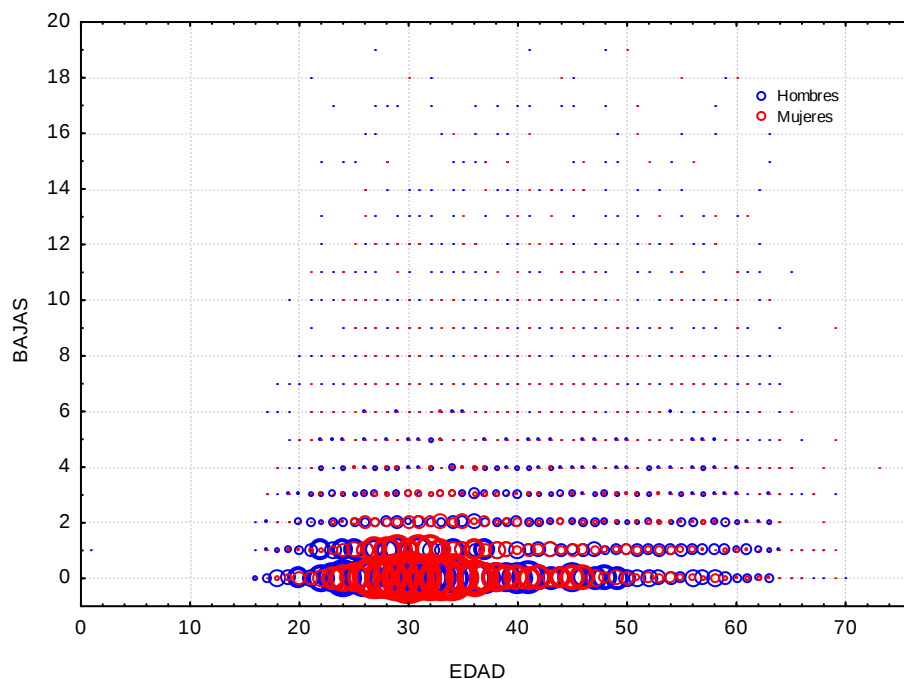
Figura 78. Frecuencias relativas del n° de bajas previas por sexo.



Nº de bajas por edad y sexo

Podemos observar en la Figura 79 que para valores de 0, 1 y 2 bajas previas las frecuencias para ambos sexos son mayores para edades comprendidas entre los 20 y los 40 años (hasta 50 para el caso de 0 bajas). Esto está en consonancia con la distribución de la muestra por edades. Mayor número de bajas es poco frecuente y se dispersa mucho en lo que a la edad se refiere. Encontramos bajas previas para ambos sexos a lo largo de todo el rango de edades.

Figura 79. Relación del número de bajas con la edad y el sexo.



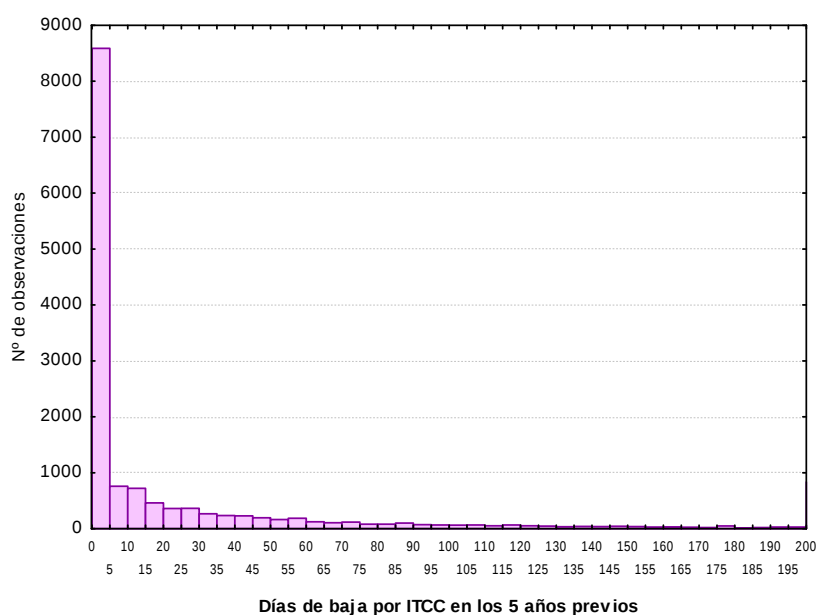
2. Número de días de baja en los 5 años previos

El 48% de los pacientes de nuestra muestra tienen días de baja por ITCC en los 5 años previos. El número de observaciones decae progresivamente a la vez que aumentan los días de baja previos (Figura 80). La mediana muestral es 0 (más del 50% de los casos no tienen días de baja) y la media muestral 40,3 días. Si consideramos solo los casos que tienen días de baja, la mediana es de 32 días y la media 84 días (Tabla 9 y Figura 80).

Tabla 9. Días de baja por ITCC en los 5 años previos.

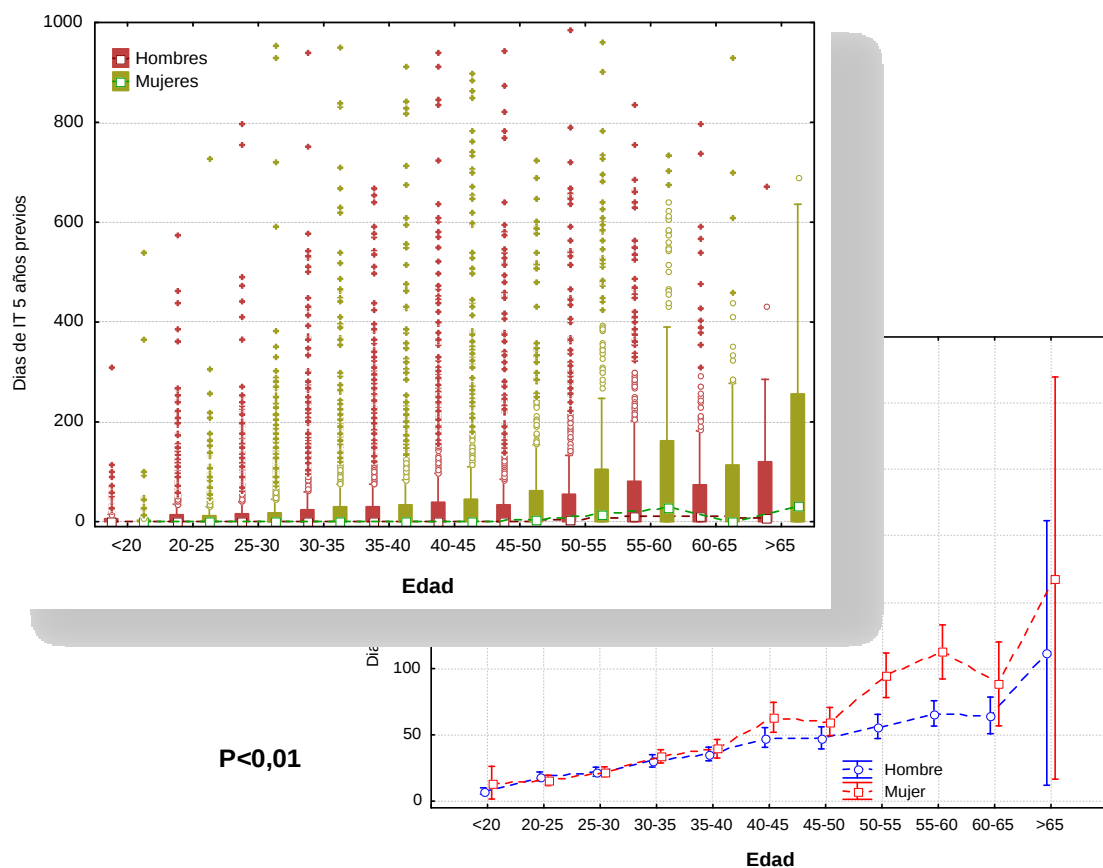
Días de baja	Muestral	Solo con bajas previas
N	14.961	7.171
IC -95%	38,7	81,1
Media	40,3	84,1
IC +95%	42,0	87,23
DE	101,2	133
Varianza	10251,4	17.689
Suma	603.479	603.479
Mínimo	0,0	1
C25%	0,0	12
Mediana	0,0	32
C75%	30,0	91
Máximo	985,0	985

Figura 80. Histograma del nº de días de baja previos.



Al analizar la relación entre el número de días de baja en los 5 años previos, la edad y el sexo, encontramos una tendencia clara a aumentar con los años. También observamos que tiende a ser mayor el número de días de baja previos en las mujeres a partir de los 40 años; diferencia que se hace mayor a partir de los 50 años. Las diferencias observadas son estadísticamente significativas ($p < 0,01$) para la mayoría de los subgrupos que pueden formarse por edad y sexo. En los siguientes gráficos (Figura 81) podemos ver las distribuciones (con medianas y cuartiles) y las medias (con sus IC al 99%) por edad y sexo.

Figura 81. Relación entre los días de baja previos, la edad y el sexo.



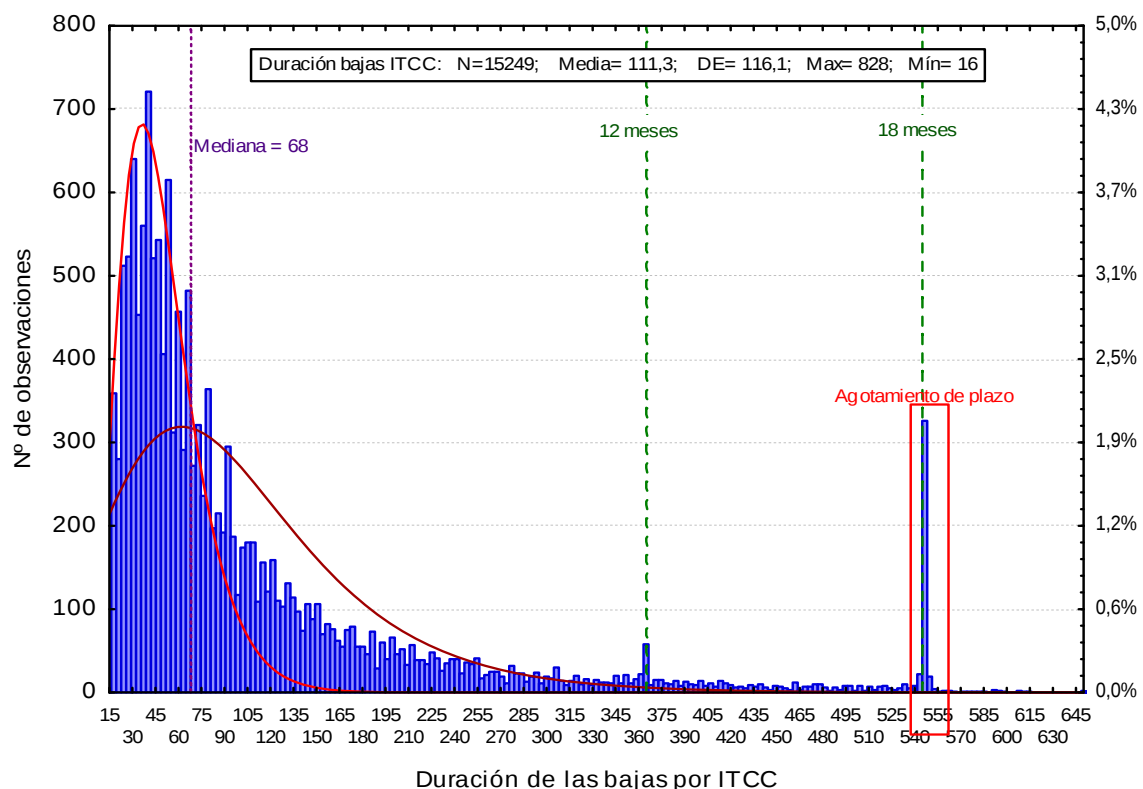
Duración de las bajas por ITCC

1. Duración de la IT

La distribución Gamma es la que mejor se aproxima a la distribución de los valores de las duraciones de la ITCC que tenemos en la muestra, si bien los resultados que obtenemos de las pruebas de bondad de ajuste (Kolmogorov-Smirnov; Chi-cuadrado) no nos permiten afirmar que los mismos se ajusten suficientemente a dicha distribución ($p < 0,01$). Dicha distribución puede calcularse a partir de la siguiente fórmula:

$$f(x) = \left\{ \frac{1}{\Gamma(c)} \right\} \times \left[\frac{x}{b} \right]^{c-1} \times e^{-\frac{x}{b}} \quad \text{para } x \geq 0, b=34 \text{ y } c=2.$$

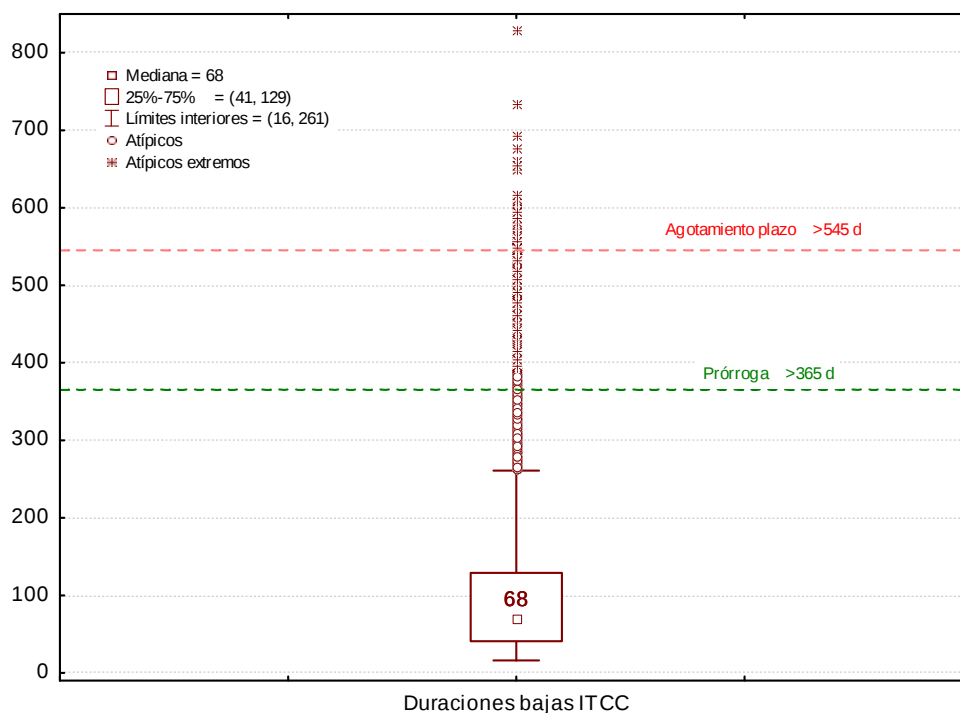
Figura 82. Distribución de la duración de las bajas por ITCC.



Por lo tanto, teniendo en cuenta dicha distribución, que presenta una gran asimetría positiva (con una cola muy larga por la derecha), lo más recomendable, en el momento de describirla, parece ser el utilizar la mediana como estimador de la

tendencia central y los cuartiles (y otros percentiles, en su caso) como estimaciones de la dispersión de los valores para la duración de la IT. Podemos fijarnos, también, en la diferencia, muy marcada, entre los valores calculados para la mediana (68) y para la media (111,3). Son, además, muy llamativas dos elevaciones de la frecuencia absoluta de valores de la duración de la ITCC que coinciden con 365 días y, muy especialmente, los que coinciden con 545 días (Figura 82). Éstos últimos son los casos que se encuentran en “agotamiento de plazo” como veremos más adelante. Podemos ver la distribución de la duración de la IT con su mediana y los cuartiles en la Figura 83. También podemos contemplar en la misma figura los valores atípicos (entre 261 y 393 días) y extremos (a partir de 394 días). Podemos comprobar en dicho diagrama la distribución asimétrica de los valores y la presencia de numerosos valores atípicos que pueden alcanzar cifras muy elevadas y, por ello, pueden influir marcadamente sobre el valor de la media, aumentándolo. En el gráfico están señalados con sendas líneas dos límites, a saber: el que se corresponde con 365 días (línea verde) que marca el límite entre periodo sin prórroga y el inicio de la misma; y el correspondiente a los 545 días (línea roja) que marca el “agotamiento de plazo”. Los casos por encima de ambos límites se corresponden con casos extremos.

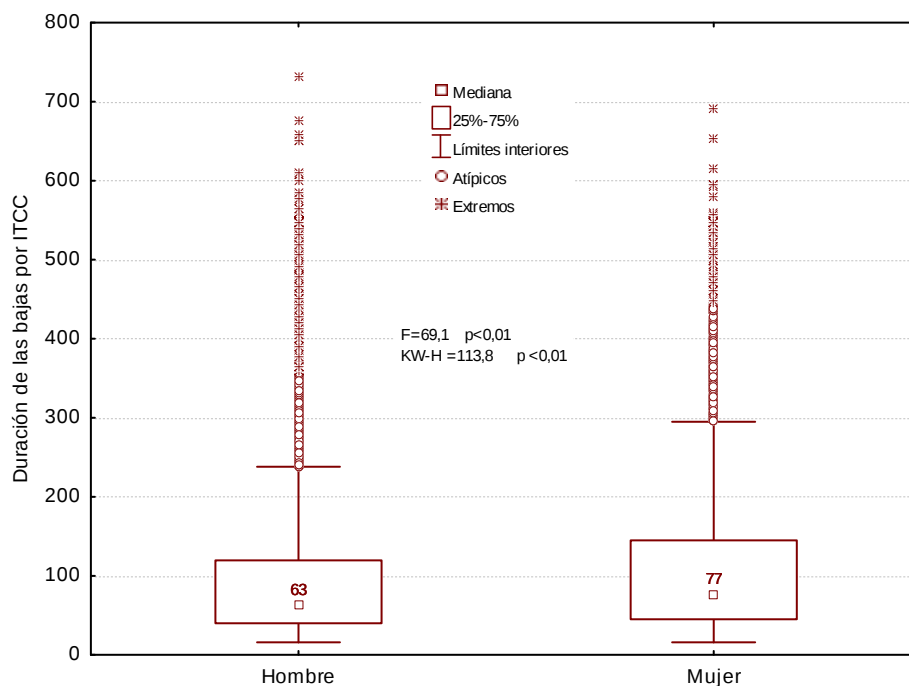
Figura 83. Duración de las bajas por ITCC.



Duración de la IT por sexo

Las distribuciones de la duración de la IT son diferentes para cada sexo. Las duraciones medias son de 106 días para los hombres y 122 días para las mujeres; además para éstas últimas encontramos una mayor variabilidad ($DE=123,3$ frente a $DE=113,2$ para los hombres); y dichas diferencias son estadísticamente significativas ($F=69,1$; $KW-H=113,8$; $p<0,01$ para ambas pruebas). Como podemos observar en la Figura 84, encontramos una mayor dispersión de los valores de la variable dependiente *duración de la IT* para la categoría “mujer”; además, para dicha categoría los valores tienden a ser más elevados que para “hombre”. Las medianas, así como los cuartiles y el límite interior superior, presentan valores más elevados para “mujer”. Para ambas categorías hay abundancia de valores atípicos que se corresponden con duraciones especialmente largas y, en muchos casos, con bajas en situación de “prórroga” (más de 12 meses) o de “agotamiento de plazo” (más de 18 meses).

Figura 84. Duración de las bajas por ITCC comparando hombres y mujeres.

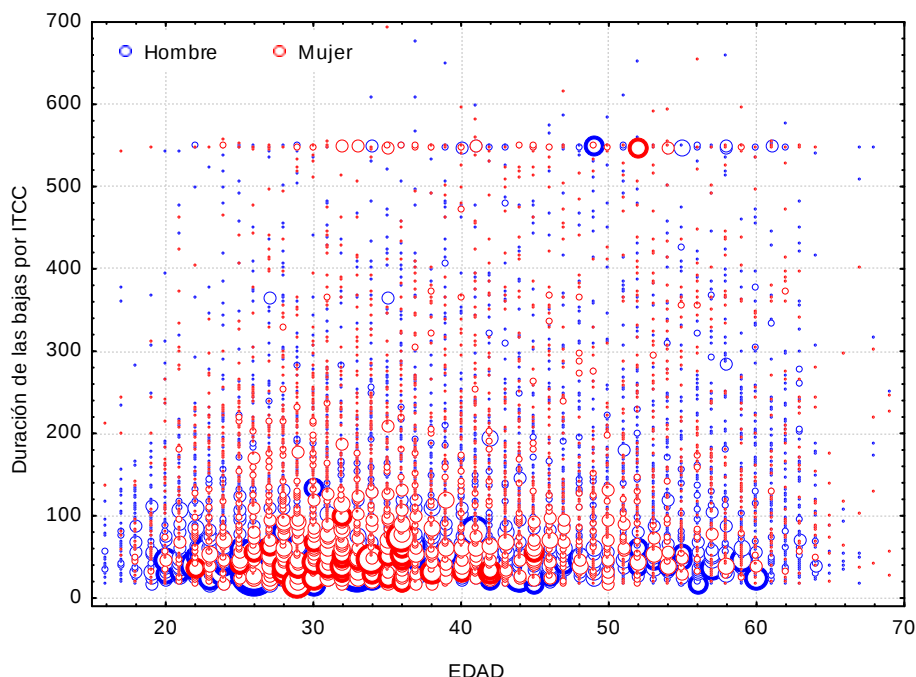


Duración de la IT por edad

En la Figura 85 podemos observar que para la mayoría de los casos de nuestra muestra las bajas por ITCC han tenido duraciones inferiores a los 100 días y la gran mayoría han durado menos de 1 año para ambos sexos y para todo el rango de edades. También podemos observar que ha habido una gran variabilidad en las

duraciones: podemos encontrar casos para todo el rango de valores de la *duración de la IT* hasta superar, incluso, el agotamiento de plazo y para todo el rango de valores de *edad*. Destaca un grupo de valores en torno a los 545 días, que se corresponden con los pacientes en agotamiento de plazo.

Figura 85. Relación de las duraciones de las bajas con la edad y el sexo.

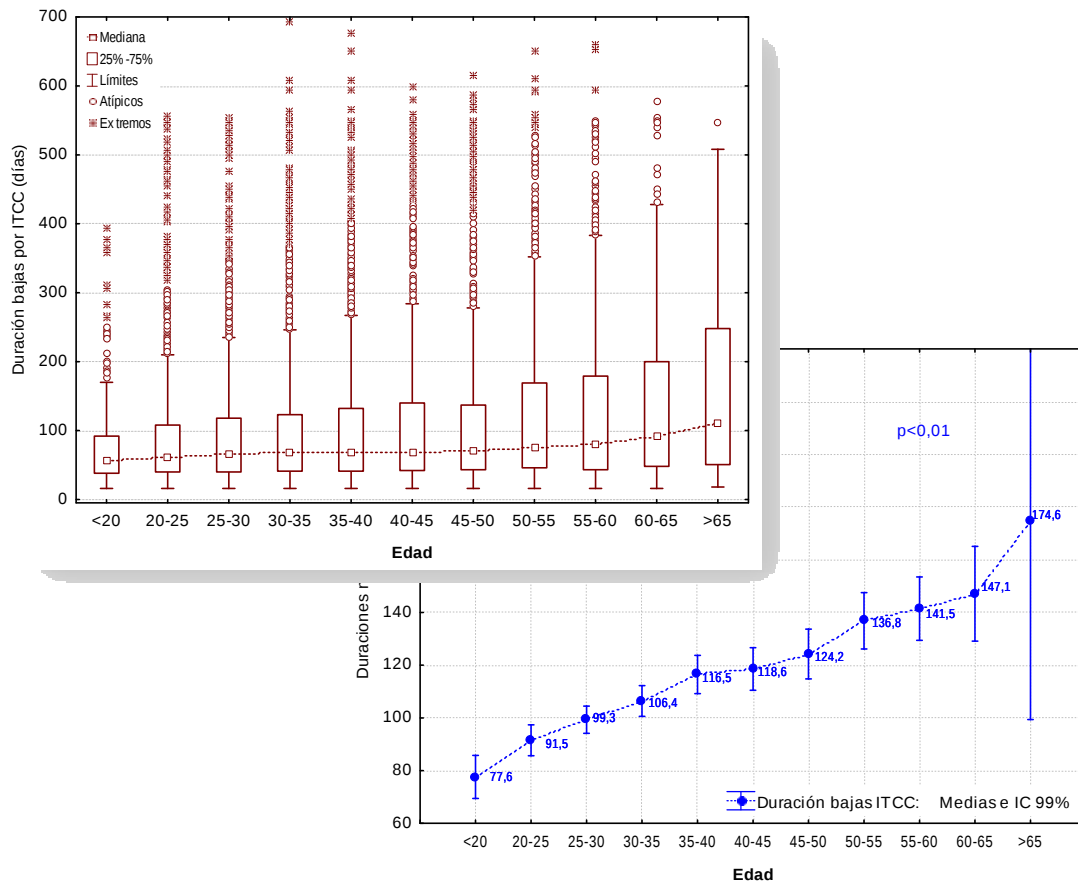


Por otro lado, si analizamos más a fondo estas relaciones, encontramos que las duraciones medias de la baja por ITCC aumentan progresivamente a la vez que lo hace la edad (tanto si tomamos ésta como intervalos o como variable continua). De hecho, encontramos una relación lineal, con una correlación positiva, entre ambas variables aunque débil debido a la gran dispersión de los valores ($p < 0,01$; $R = 0,14$; $R^2 = 0,02$). También podemos observar en la Figura 86 cómo aumentan las medianas, los cuartiles y el rango intercuartil a la vez que lo hace la edad. Es decir, con edades más altas encontramos con más frecuencia duraciones más largas y una mayor dispersión en los valores. Las diferencias observadas en la duración de la IT entre los diferentes intervalos de edad son estadísticamente significativas ($F = 29,8$; $p < 0,01$; K-W $H = 176,2$; $p < 0,01$). Las medias y las medianas de la duración de las bajas, para cada uno de los intervalos de edad, siguen un crecimiento lineal; de hecho, se ajustan bastante bien a sendas rectas de regresión con las siguientes fórmulas:

$$\text{Duración media de la baja} = 72,06 + 8,20 \times \text{intervalo de edad (1, 2...)}$$

Mediana de la duración = $49,81 + 4,07 \times \text{intervalo de edad (1, 2...)}$

Figura 86. Duración de las bajas por ITCC y su relación con intervalos de edad.

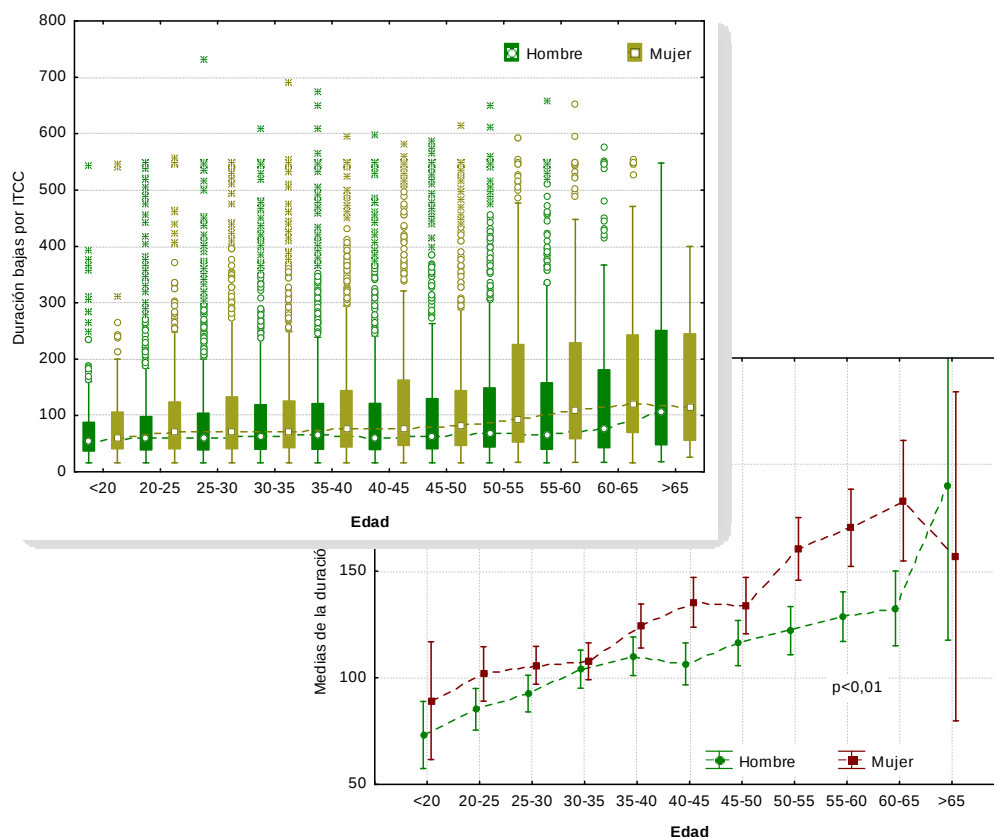


Duración de la IT por edad y sexo

Si repetimos el análisis anterior introduciendo en el modelo, además, el sexo, encontramos que, como ya hemos visto, las duraciones de las bajas por ITCC aumentan a la vez que lo hace la edad y, además, tienden a ser más largas para las mujeres para todos los intervalos de edad (Figura 87). Las diferencias observadas son estadísticamente significativas ($p < 0,01$). Una excepción sería el intervalo de edad para mayores de 65 años, pero en este caso tenemos muy pocos datos (17 hombres y 16 mujeres) y las diferencias no son estadísticamente significativas. Todos los efectos sobre la duración de la ITCC (del sexo, de la edad y de la interacción entre ambas) son estadísticamente significativos ($p < 0,01$), pero son muy débiles (pueden explicar aproximadamente el 0,1%, 2,2% y 0,2% de la variabilidad de duración de la bajas

respectivamente). Si realizamos los análisis tomando la edad como variable continua (ANCOVA) encontramos resultados consistentes con lo ya visto. Efectos estadísticamente significativos ($p < 0,01$) pero débiles tanto para el sexo, como para la edad y la interacción entre ambas variables. El efecto más importante, como ya hemos visto, es el que corresponde a la edad.

Figura 87. Duración de las bajas por ITCC y su relación con intervalos de edad y con el sexo.



También se observan distribuciones diferentes para la duración de la IT en relación con dichas variables. Al igual que sucedía con las medias, las medianas y los rangos intercuartiles tienden a aumentar con la edad y con el sexo femenino. Los rangos intercuartiles también tienden a aumentar con la edad.

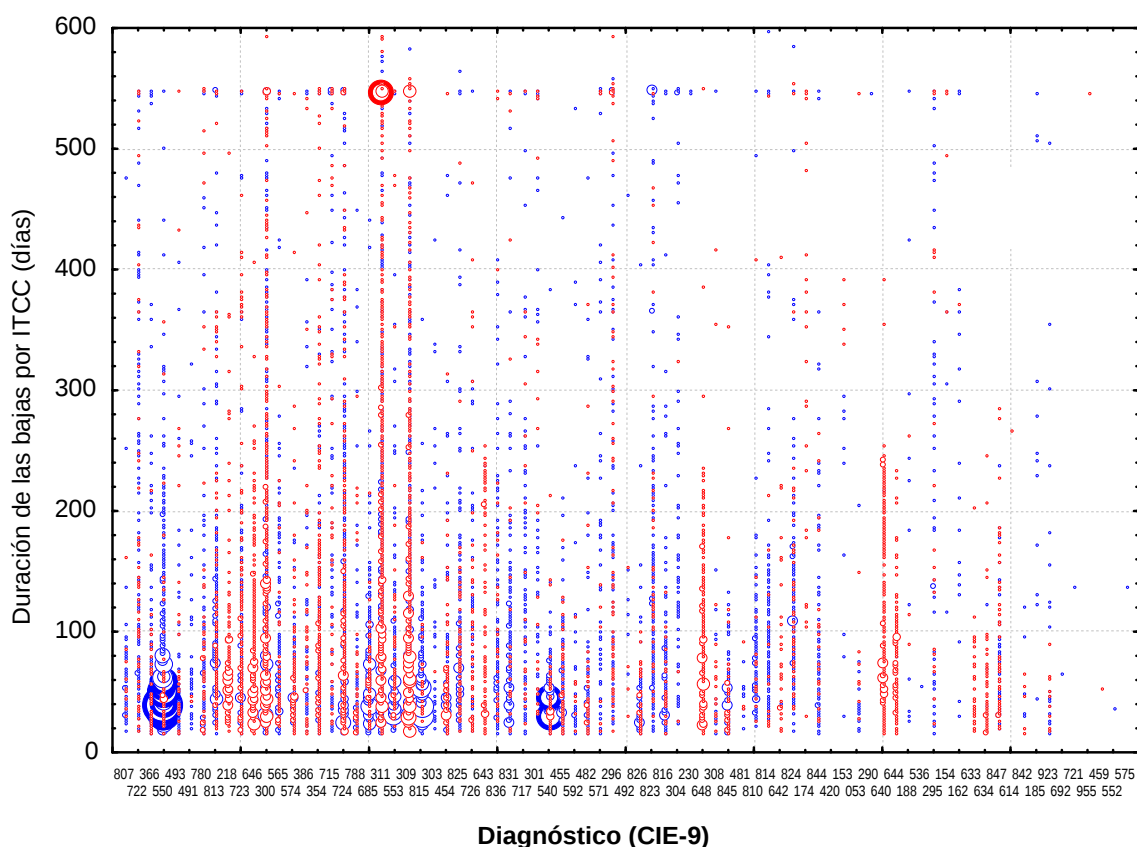
A pesar de que encontramos un efecto tan evidente, aunque débil, de las variables *sexo* y *edad* sobre la duración de las bajas por ITCC, debemos de considerar que, si bien ambas parecen tener alguna influencia independiente sobre la duración de la IT, es muy probable que actúen, sobre todo, como variables de confusión. Podemos observar en diversas variables que también influyen sobre la duración de la IT

diferencias en sus valores relacionadas con el sexo y la edad. Un ejemplo de dichas variables pueden ser: el nivel de estudios, la ocupación, la actividad laboral y, especialmente, el diagnóstico, entre otras.

Duración de la IT por diagnóstico

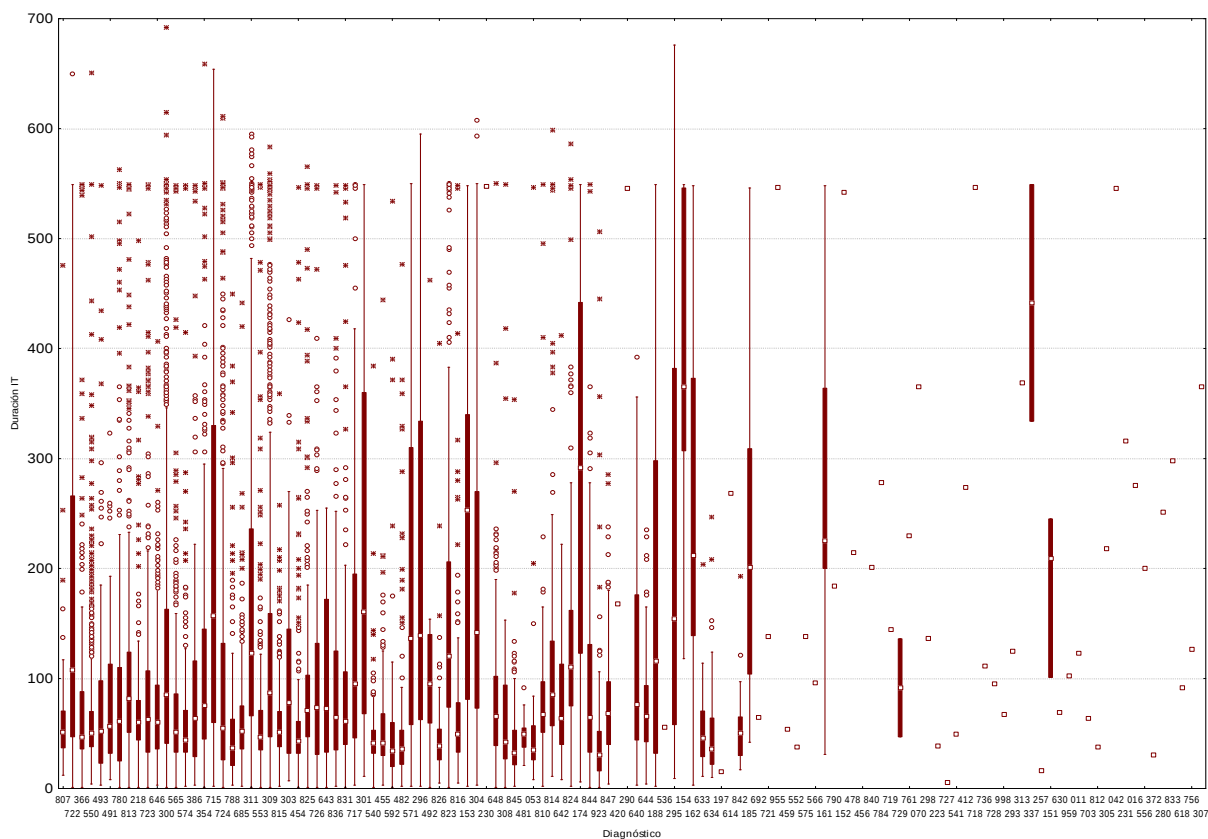
Simplemente dedicando una ojeada rápida a la Figura 88 podemos concluir que hay diferencias marcadas en cuanto a las duraciones de las bajas para los diferentes diagnósticos, como muestra también la experiencia clínica. Además podemos observar las relaciones de ambas variables con el sexo. Dado que hay diferencias en las frecuencias de los procesos patológicos diagnosticados para cada uno de los géneros y que hay diferentes duraciones para cada uno de los procesos patológicos, una consecuencia lógica de todo ello es que haya diferencias en las duraciones de la ITCC por sexo. Como veíamos anteriormente, no parece que el sexo, por sí mismo, pueda ser factor importante por su influencia sobre la duración de las bajas por ITCC sino, más bien, un factor de confusión; una variable relacionada con otras que sí pueden tener una influencia más directa sobre la misma, como es el caso del diagnóstico.

Figura 88. Duraciones de las bajas por ITCC para cada diagnóstico (CIE-9) categorizadas por sexo.



En la Figura 89 podemos observar las diferentes distribuciones de las duraciones para cada uno de los diagnósticos. Podemos apreciar una gran variabilidad en las medianas, cuartiles y rangos intercuartiles. Es decir, encontramos parámetros de tendencia central y dispersiones de los datos muy diversas. Para unos diagnósticos las duraciones pueden alargarse mucho, mientras que para otros suelen ser mucho más cortas. Para algunos procesos puede encontrarse gran variedad de duraciones, para diferentes pacientes, con grandes diferencias entre unas y otras; mientras que para otros procesos la duración está más acotada y varía poco de unos a otros pacientes. También vemos un número variable de casos atípicos para bastantes diagnósticos.

Figura 89. Duraciones de las bajas por ITCC para cada diagnóstico.

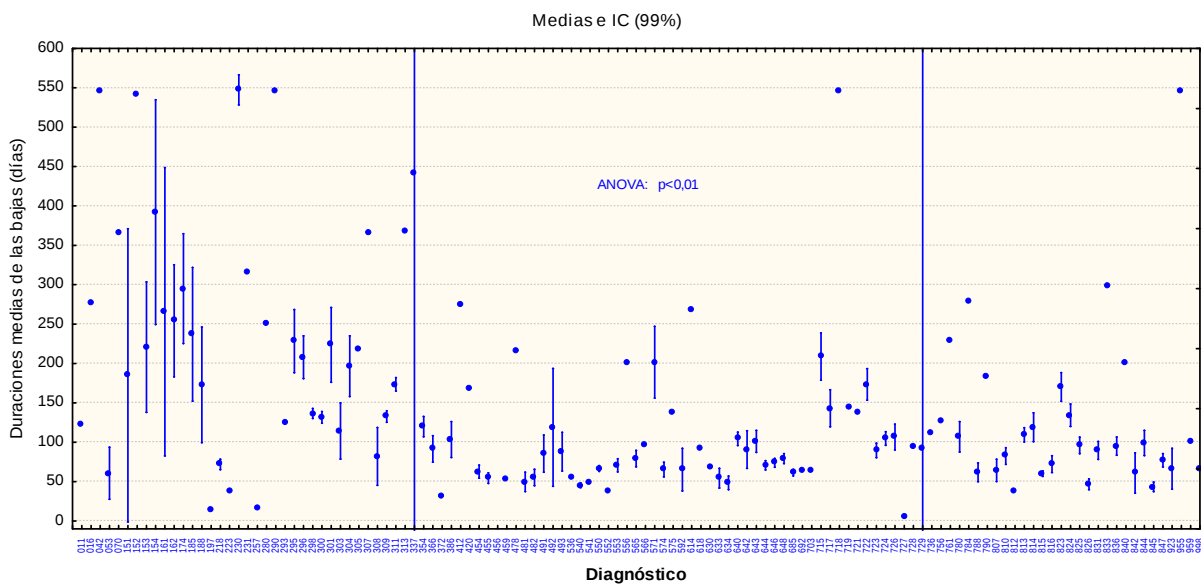


En definitiva, existe una gran diversidad entre los diferentes diagnósticos en cuanto a la duración del proceso y de la baja laboral asociada al mismo. Sin embargo podemos observar que hay grupos de diagnósticos que se parecen entre sí en lo que a las distribuciones de dichas duraciones se refiere. Ello nos permite agruparlos en unas pocas categorías, aunque no se parezcan entre sí, por las características clínicas propias del proceso diagnosticado, como ya vimos en la página 63 (

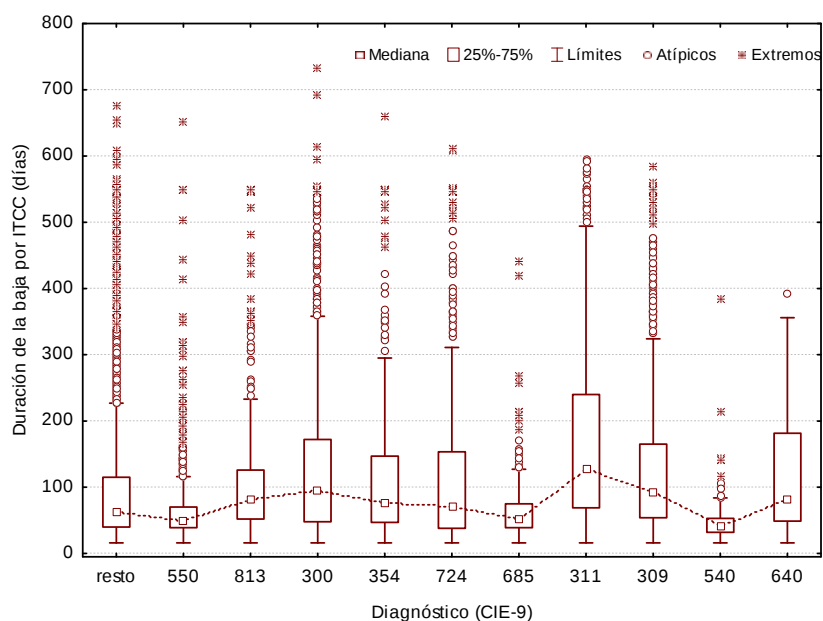
Reducción del nº de categorías de las variables Diagnóstico y Actividad económica).

Lo mismo podemos observar en un gráfico que represente a las duraciones medias (Figura 90) para cada uno de los diagnósticos. Las diferencias observadas entre dichas medias son estadísticamente significativas ($F=26,27$; $p<0,01$; K-W $H=2782,5$; $p<0,01$), aunque no entre todos los diagnósticos (de dos en dos).

Figura 90. Medias e IC al 99% de las duraciones de las bajas laborales.



En la Figura 91 mostramos las distribuciones de las duraciones de las bajas laborales para los 10 diagnósticos que serán analizados independientemente en este estudio.

Figura 91. Duraciones de las bajas laborales para los 10 diagnósticos seleccionados.

Podemos destacar las diferencias fáciles de observar entre diagnósticos como el 540, el 685 y el 550 y otros como el 311, el 300 y 309. Los primeros son de carácter quirúrgico, más fáciles de objetivar en la clínica y con procesos, generalmente, más cortos, con tratamientos y resultados más definidos, menos complicados y más fáciles de controlar. El segundo grupo se corresponde con procesos mentales más difíciles de objetivar en la clínica y de controlar; generalmente son más prolongados y más complicados de tratar.

El diagnóstico es un factor que tiene una gran influencia sobre la duración de la ITCC. El efecto es estadísticamente significativo ($p < 0,01$) para el modelo en conjunto. Sin embargo, las diferencias observadas no siempre son estadísticamente significativas cuando se comparan los diagnósticos dos a dos. Nuevamente, encontramos lo que ya hemos indicado anteriormente: se trata de un factor con un efecto importante y, aparentemente, directo sobre la duración de las bajas laborales; y además, hay diagnósticos que son similares por su efecto sobre la duración de la IT, independientemente de sus características clínicas. Cuando los agrupamos con este criterio (ver página 63 y Anexo 8), si encontramos diferencias estadísticamente significativas ($p < 0,01$) entre todos los grupos.

El lector podrá encontrar más información relacionada con este tema en los 10 documentos independientes dedicados al análisis de la IT para los 10 diagnósticos

mencionados, así como en el apartado que trata de la comparación de la IT entre los diagnósticos 550 y 311 de la página 152 (titulado “Comparaciones entre dos diagnósticos”).

Duración IT y comorbilidad

Otro factor relevante que tiene un efecto estadísticamente significativo ($p < 0,01$) sobre las duraciones de las bajas laborales por ITCC en nuestra muestra es la *comorbilidad*. En las figuras 92 y 93 pueden observarse las diferencias en las distribuciones y en los parámetros estimados para el grupo de pacientes con comorbilidad y el grupo que no la presenta. Las duraciones tienden a ser ostensiblemente más largas en los pacientes del grupo con comorbilidad. Así, la media para el grupo con comorbilidad es de 167,5 días frente a 97 días en el grupo sin comorbilidad. Las medianas para ambos grupos son 109 para el primero frente a 60 para el segundo. La dispersión de los valores es mucho mayor para el grupo con comorbilidad, 58 días (Q_1) y 232 (Q_3), que para el grupo sin comorbilidad, 36 días (Q_1) y 111 (Q_3).

Figura 92. Efecto de la comorbilidad sobre la duración de las bajas laborales.

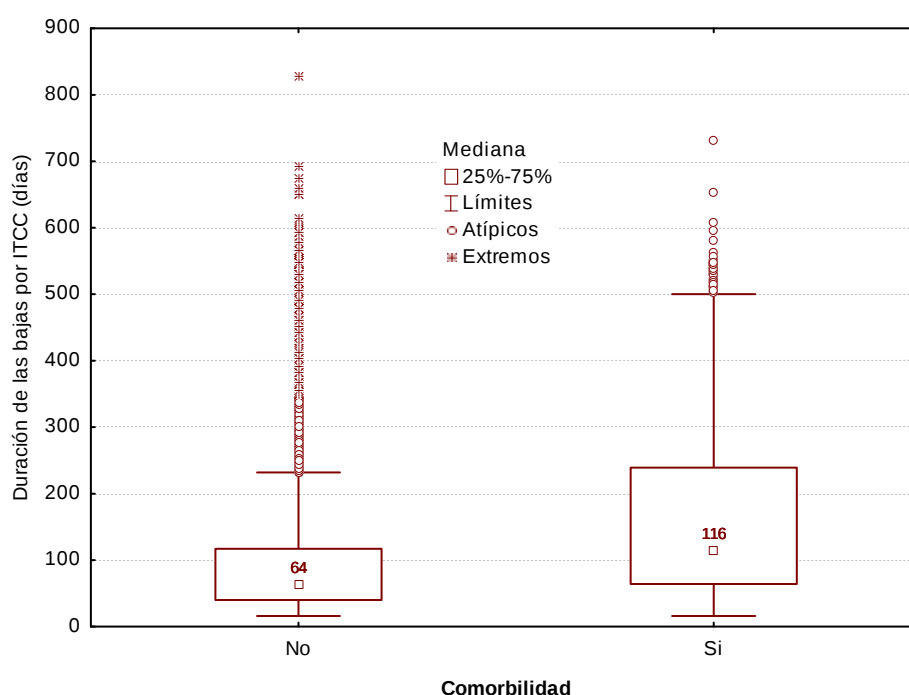
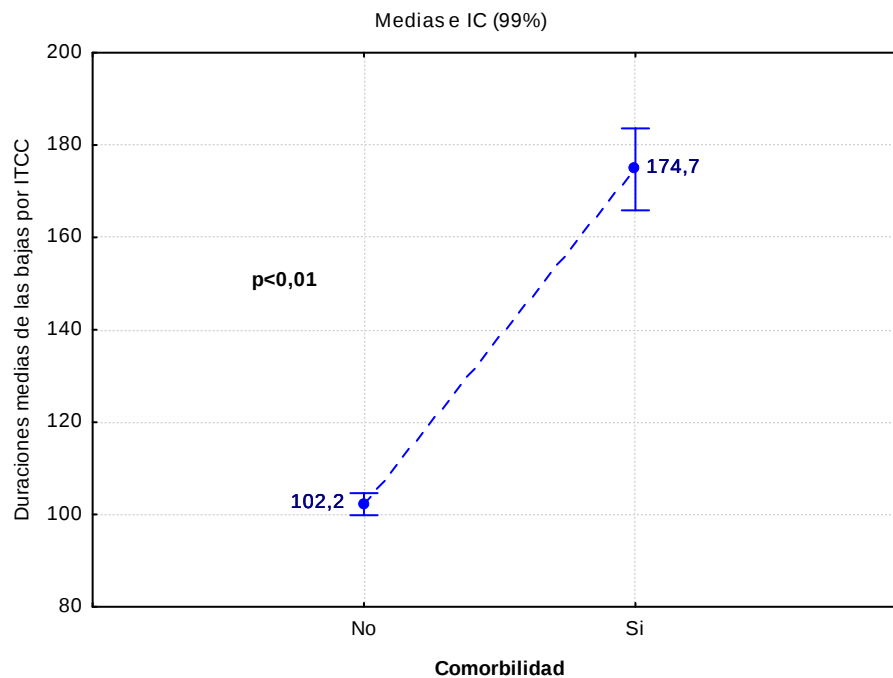


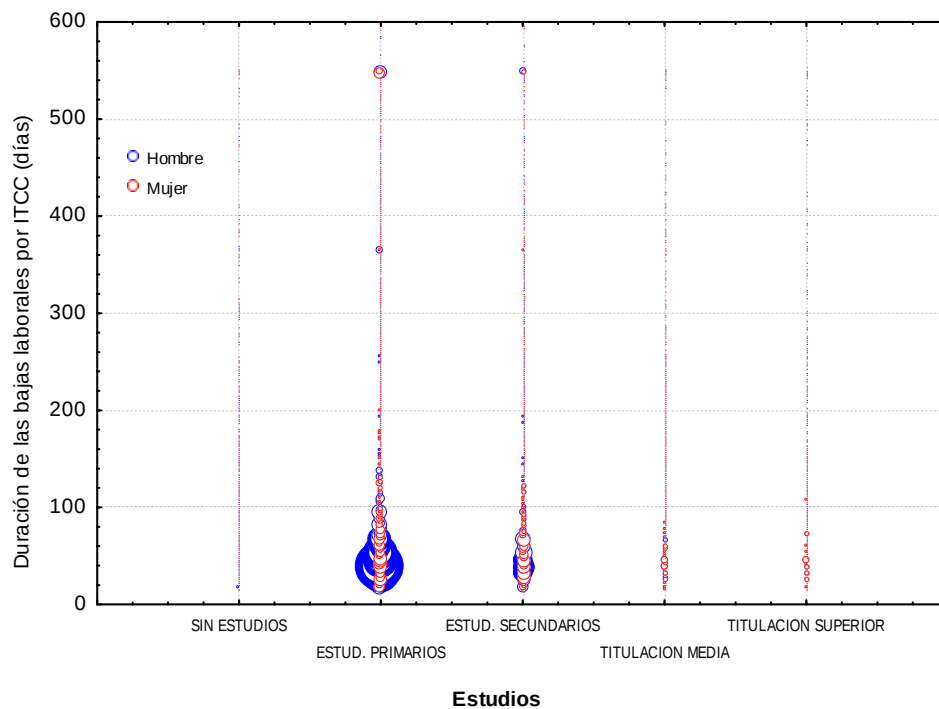
Figura 93. Diferencias en las duraciones medias de las bajas laborales con ausencia y presencia de comorbilidad



Duración IT y nivel de estudios

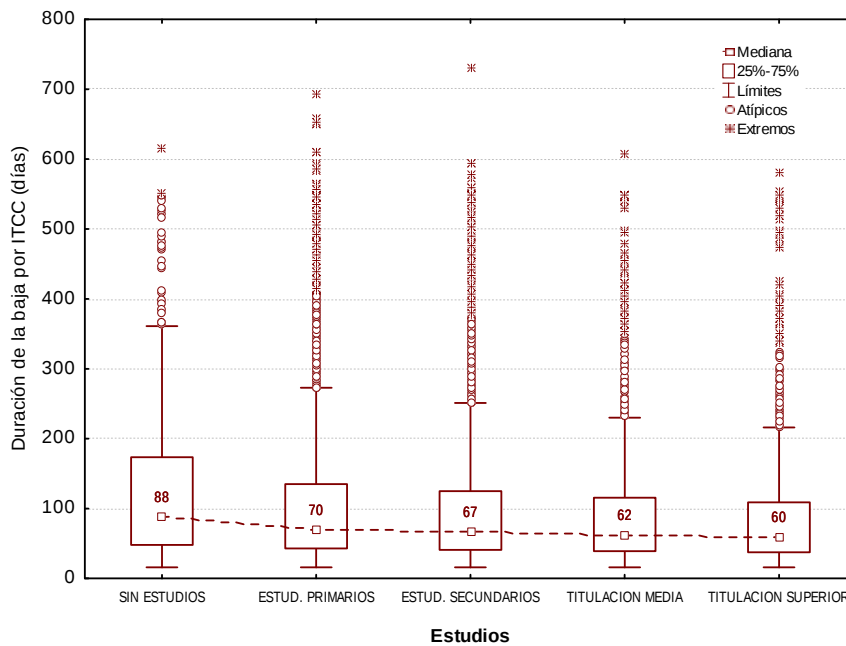
La duración de las bajas laborales por ITCC tiende a disminuir a la vez que aumenta el nivel de estudios de los pacientes. Sin embargo, deberemos ser cautelosos con este efecto ya que su origen puede estar en otros factores relacionados con el nivel de estudios como pueden ser: el puesto de trabajo, el sector y la actividad económica de la empresa, condiciones del entorno socio-laboral y familiar del paciente, etc. En la Figura 94 podemos ver representadas la dispersión de datos y las frecuencias para cada valor de la duración para cada nivel de estudios y ambos géneros.

Figura 94. Frecuencias de las duraciones de las bajas laborales por nivel de estudios.



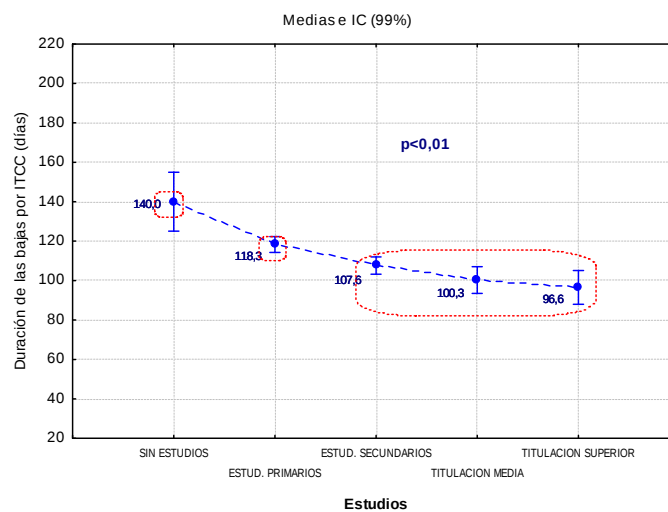
Podemos observar que las duraciones tienden a ser, en conjunto, más cortas y con menor dispersión de los valores según aumenta el nivel de estudios (Figura 95); aunque en todos los niveles puedan darse casos con duración larga. Como hemos comentado más arriba, debemos de ser cautelosos con la interpretación de estos resultados. Es evidente que hay una relación entre el nivel de estudios y la duración de las bajas, pero a su vez ambas pueden estar asociadas a otros factores de tipo laboral, socio-económico, familiar, tipo de patología, etc. Esto significa que no es fácil establecer una relación del tipo causa-efecto directa pero, en cambio, sí puede ser muy útil como marcador de pacientes que pueden tener bajas más largas. De alguna forma puede ser una variable que englobe los efectos de esos otros factores que comentábamos.

Figura 95. Distribuciones de las duraciones de las bajas laborales por niveles de estudios.



Realizando un análisis de las varianzas, o bien con pruebas no paramétricas, encontramos diferencias estadísticamente significativas ($F=22,2$; $p<0,01$; K-W $H=90,6$; $p<0,01$) entre tres grupos: trabajadores sin estudios, trabajadores con estudios primarios y el conjunto formado por los trabajadores con estudios secundarios o con titulaciones medias o superiores. Entre estos tres, que forman el tercer grupo, no encontramos diferencias estadísticamente significativas (Figura 96).

Figura 96. Duraciones medias de las bajas para cada nivel de estudios.



Las duraciones medias para cada uno de los grupos, así como las medianas correspondientes, se ajustan bastante bien a sendas rectas definida por las siguientes fórmulas:

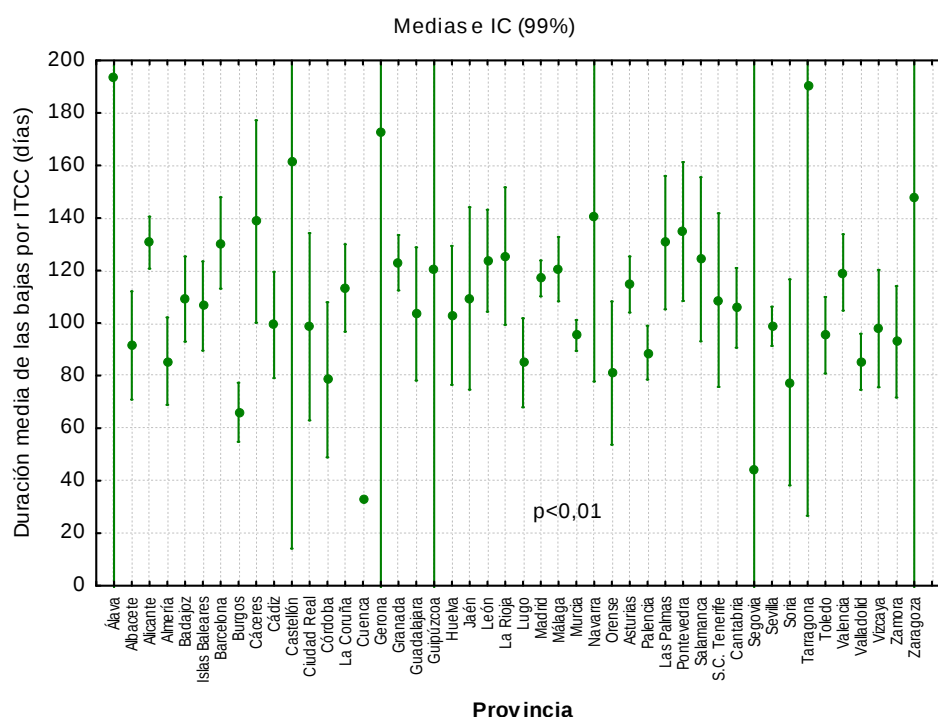
$$\text{Duración media de la baja} = 144,03 - 10,5 \times \text{Nivel de estudios (0, 1, 2...)}$$

$$\text{Mediana de la duración de la baja} = 88,6 - 6,4 \times \text{Nivel de estudios (0, 1, 2...)}$$

Duración IT por provincia

Hay diferencias en la duración de la IT por provincias (Figura 97), sin embargo, aunque el modelo en su conjunto es estadísticamente significativo ($p < 0,01$) encontramos diferencias solo entre algunas provincias. Además en algunas de ellas tenemos muestras muy pequeñas. Por último, dichas diferencias pueden deberse a varios factores que pueden ser independientes, o no, de la provincia donde recibió la baja. Así, por ejemplo, podemos encontrar diferentes prevalencias para los distintos diagnósticos por provincia y dichos procesos pueden tener duraciones muy dispares.

Figura 97. Duraciones medias de las bajas por provincia.



En la siguiente tabla podemos consultar las medias y desviaciones de las duraciones de las bajas laborales para cada una de las provincias (Tabla 10).

Tabla 10. Medias y DE de las duraciones por provincia.

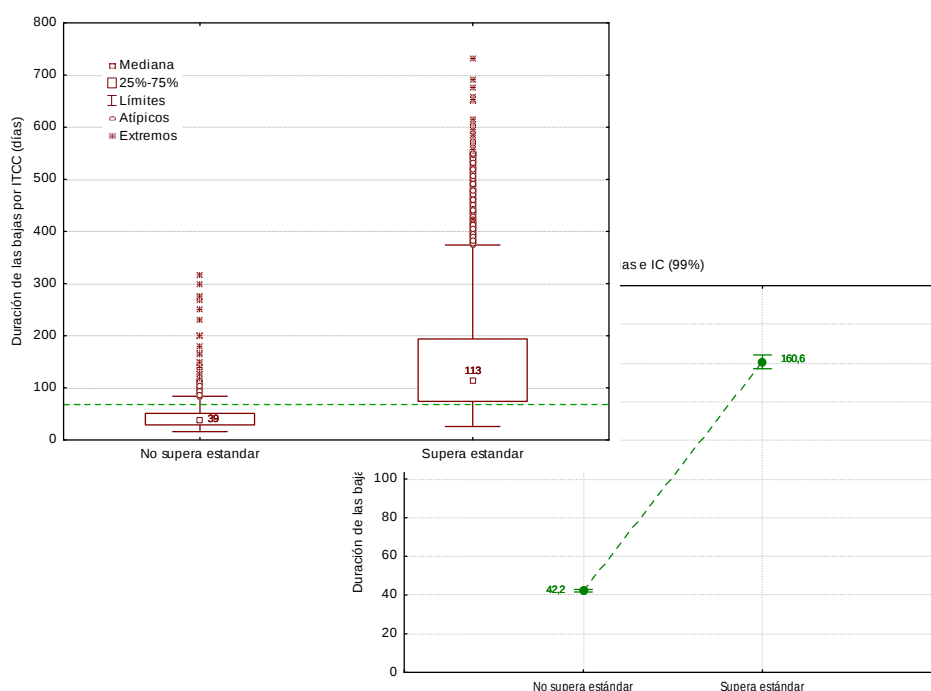
	N	Medias	IC -99%	IC +99%	DE
Álava	7	193,86	-152,49	540,21	247,17
Albacete	204	91,49	70,92	112,06	113,01
Alicante	1190	130,63	120,73	140,53	132,37
Almería	146	85,53	68,91	102,16	76,95
Badajoz	314	109,16	92,95	125,36	110,77
Islas Baleares	220	106,55	89,56	123,53	96,96
Barcelona	384	130,53	113,13	147,93	131,69
Burgos	170	66,03	54,79	77,27	56,26
Cáceres	111	138,68	100,15	177,21	154,86
Cádiz	151	99,32	79,10	119,53	95,20
Castellón	11	161,27	14,16	308,38	153,95
Ciudad Real	56	98,64	63,01	134,28	99,93
Córdoba	40	78,43	48,91	107,94	68,93
La Coruña	325	113,38	96,75	130,02	115,76
Cuenca	1	33,00			0,00
Gerona	5	173,00	-249,23	595,23	205,06
Granada	923	123,02	112,46	133,57	124,27
Guadalajara	110	103,51	78,16	128,86	101,42
Guipúzcoa	5	120,20	-111,76	352,16	112,66
Huelva	152	102,95	76,47	129,42	125,13
Jaén	79	109,42	74,68	144,15	116,93
León	256	123,75	104,35	143,14	119,60
La Rioja	166	125,53	99,35	151,71	129,42
Lugo	67	84,93	67,98	101,88	52,31
Madrid	2137	117,03	110,19	123,87	122,64
Málaga	682	120,56	108,30	132,82	123,92
Murcia	1651	95,30	89,40	101,19	92,90
Navarra	47	140,68	77,76	203,60	160,52
Orense	62	80,98	53,72	108,25	80,74
Asturias	751	114,73	104,10	125,36	112,80
Palencia	494	88,78	78,55	99,01	87,93
Las Palmas	174	130,64	105,23	156,06	128,71
Pontevedra	148	134,89	108,42	161,35	123,36
Salamanca	150	124,35	93,13	155,56	146,52
S.C. Tenerife	77	108,78	75,72	141,84	109,81
Cantabria	282	105,80	90,63	120,98	98,23
Segovia	2	44,50	-114,64	203,64	3,54
Sevilla	1064	98,76	91,29	106,23	94,45
Soria	20	77,45	38,25	116,65	61,28
Tarragona	11	190,64	26,64	354,63	171,62
Toledo	390	95,39	80,84	109,94	111,00
Valencia	590	119,32	104,81	133,84	136,43
Valladolid	495	85,30	74,63	95,97	91,81
Vizcaya	97	97,93	75,58	120,27	83,73
Zamora	140	92,88	71,66	114,09	96,12
Zaragoza	4	147,50	-302,87	597,87	154,21
Todas	14561	110,31	107,86	112,76	114,73

Casos que superan el estándar de duración de Ibermutuamur

Tomando para cada diagnóstico su propio estándar de Ibermutuamur, podemos observar en la Figura 98 cómo los casos que no superan dichos estándares tienen en conjunto duraciones de la IT mucho más cortas, obviamente, que aquellos casos que

los superan. En este caso, para simplificar los análisis, hemos reducido los diagnósticos a las categorías de la CIE-9 con tres dígitos, englobando en ellas las subcategorías correspondientes. De este modo, no solo reducimos el número de categorías sino que también evitamos recuentos muy pequeños. En el documento dedicado a los estándares de duración se analizarán de forma más pormenorizada. Prácticamente, el 90% de los casos que no superan los estándares se encuentran por debajo de la mediana muestral de la IT (68 días; línea punteada verde de la Figura 98). Aproximadamente el 80% de los casos que superan los estándares están por encima de la mediana muestral. Encontramos grandes diferencias que son estadísticamente significativas ($p < 0,01$) entre las medias de ambos grupos (160,6 días para los que superan los estándares frente a 42,2 días para los que no los superan).

Figura 98. Duraciones de las bajas laborales con relación a los estándares.



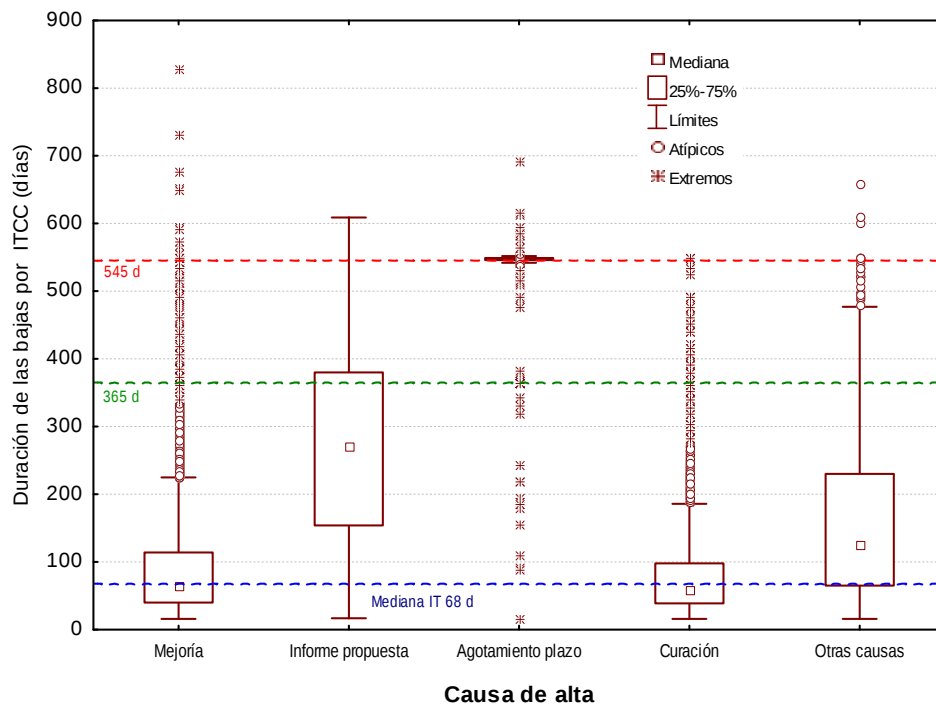
Estos resultados muestran que puede ser útil utilizar un valor de referencia que sea razonable, como los estándares de Ibermutuamur, para valorar las duraciones de los procesos y sus desviaciones con relación a lo que sería juiciosamente esperable en ausencia de otros factores, aparte del proceso mismo, como pueden ser: una evolución tórpida, mala respuesta al tratamiento, comorbilidad, etc.

Duración de la ITCC y causa del alta

Agrupando los casos en función de cuál fue la causa de alta, encontramos que los casos con “mejoría” o “curación” duran entre 16 y aproximadamente 200 días, con medianas en 64 y 59 días, respectivamente. Aunque es estadísticamente significativa ($p < 0,01$) la diferencia en la duración de la ITCC entre la “mejoría” y la “curación”, ambas se parecen mucho en cuanto a la distribución de los valores. Las dos categorías tienen una mediana de la duración de la ITCC próxima a 68 días (que es la mediana para el conjunto de la muestra) y aproximadamente el 75% de los casos están por debajo de los 100 días de duración de la IT. El rango intercuartil para ambas es bastante más corto que para el resto (a excepción de los agotamientos de plazo). Para estas dos categorías hay casos con duraciones atípicamente prolongadas. Para la categoría “otras causas” las duraciones tienden a ser más largas pero mayoritariamente dentro de los 12 meses. Solo algunos casos requieren “prórroga”. Las duraciones tienden a ser mayores, aún, para aquellos pacientes que reciben el “informe propuesta”. En estos casos la duración de la baja por ITCC se alarga mucho y presenta mayor dispersión en sus valores. Para los pacientes que alcanzan el “agotamiento de plazo” las duraciones son superiores a los 18 meses, aunque en este caso hay muy poca dispersión (debemos de recordar que algunos casos presentan una duración de la IT corta debido a que la baja actual se ha acumulado con otras anteriores por el mismo proceso hasta alcanzar el agotamiento de plazo; en nuestros datos tenemos la duración del episodio actual). Ver Figura 99.

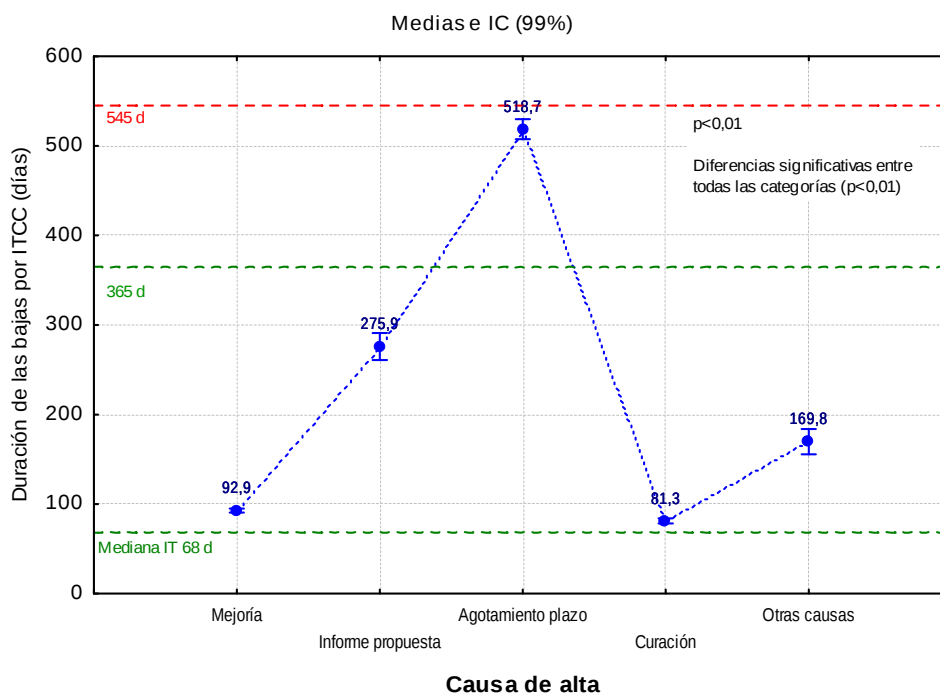
El grupo con “Otras causas” de alta se encuentran en una posición intermedia, en cuanto a la duración de la ITCC, entre “mejoría” y “curación” e “informe propuesta”.

Figura 99. Relación entre las duraciones de la IT y la causa de alta.



Encontramos para los 5 grupos diferencias estadísticamente significativas ($p < 0,01$) entre sus medias (ver Figura 100).

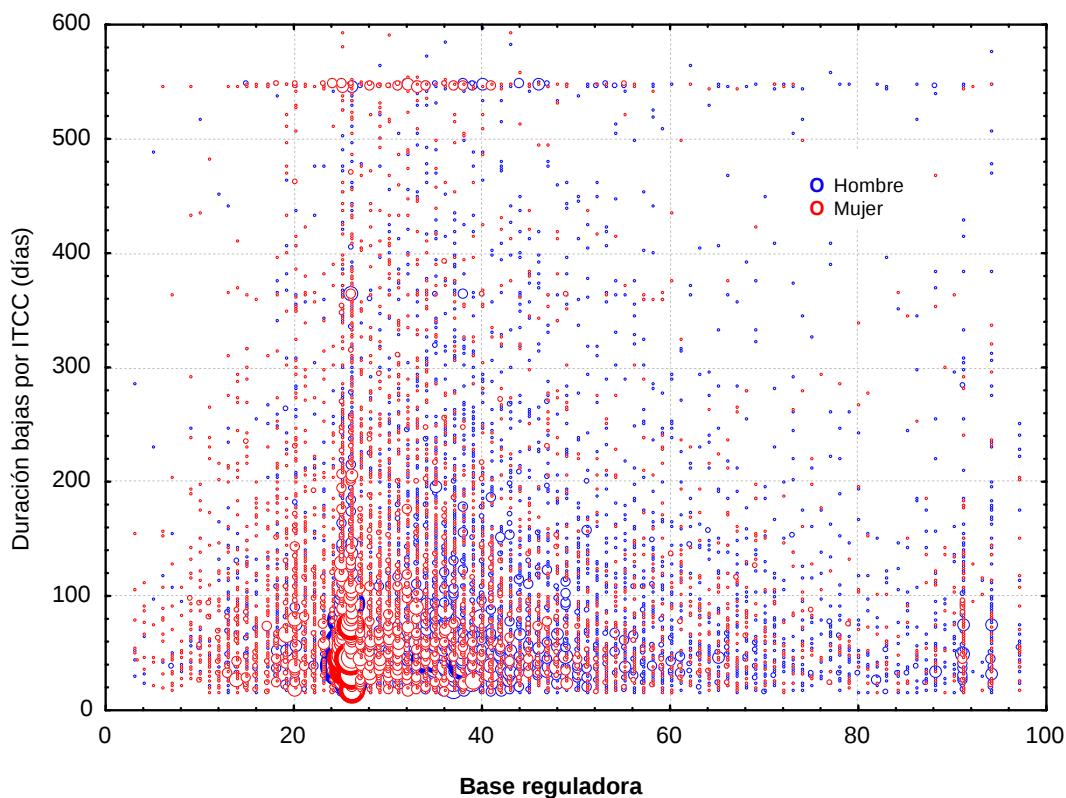
Figura 100. Duraciones medias de la IT por causa de alta.



Duración de la IT por base reguladora y sexo

Las mayores frecuencias de casos se observan, para ambos sexos, en niveles retributivos bajos o medios (20€ a 50€) y duraciones de la ITCC inferiores a 100 días (Figura 101). Los casos con duraciones más largas tienden a coincidir con dichos niveles retributivos. También la mayoría de los casos en agotamiento de plazo. En dichos niveles retributivos están, aproximadamente, el 75% de los trabajadores de nuestra muestra. En los casos que tienen bases reguladoras más elevadas, la ITCC no suele durar mucho (100 días o menos para la mayoría de los casos). Hay una correlación estadísticamente significativa, pero muy débil y negativa, entre la base reguladora y la duración de la IT ($r=-0,03$; $p<0,01$). Dicha correlación disminuye al ajustar por sexo ($r=-0,02$;). Es decir, en general observamos cierta tendencia en el sentido de que disminuyan las duraciones de las bajas a la vez que aumenta la base reguladora; pero como existe mucha dispersión de los valores y, además, disminuye mucho el número de observaciones al aumentar la base reguladora, no parece razonable extraer conclusiones definitivas de esta observación. Además, la base reguladora está relacionada con otras variables que se asocian, a su vez, con las duraciones de las bajas, como la *ocupación*, el *nivel de estudios*, etc.

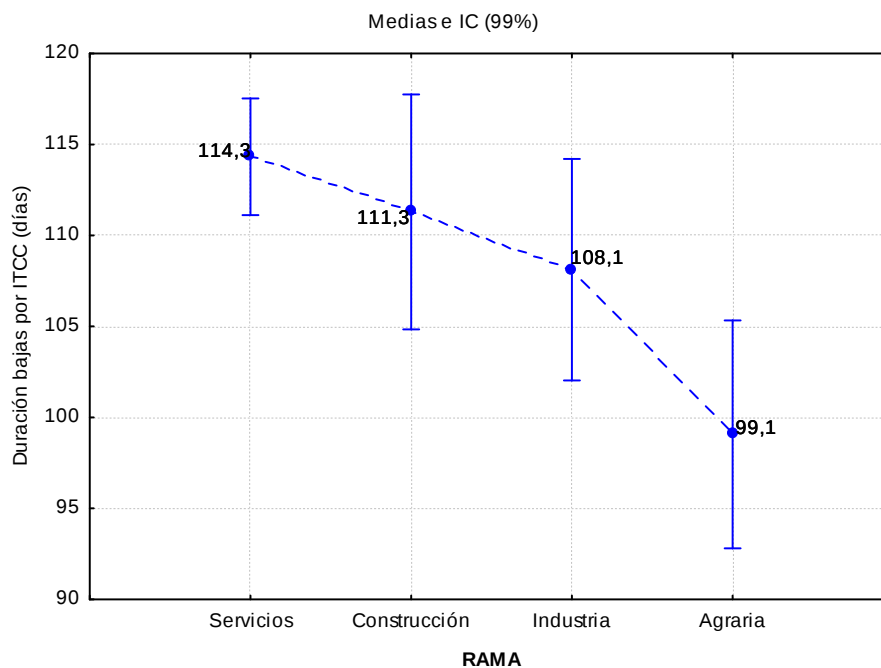
Figura 101. Relaciones entre la base reguladora, las duraciones de las bajas laborales y el sexo.



Duración de la ITCC y rama de actividad de la empresa

Los sectores Servicios y Construcción se diferencian de forma estadísticamente significativa del sector Agrario ($p < 0,01$). No existen otras diferencias estadísticamente significativas entre los diferentes sectores. Ver Figura 102.

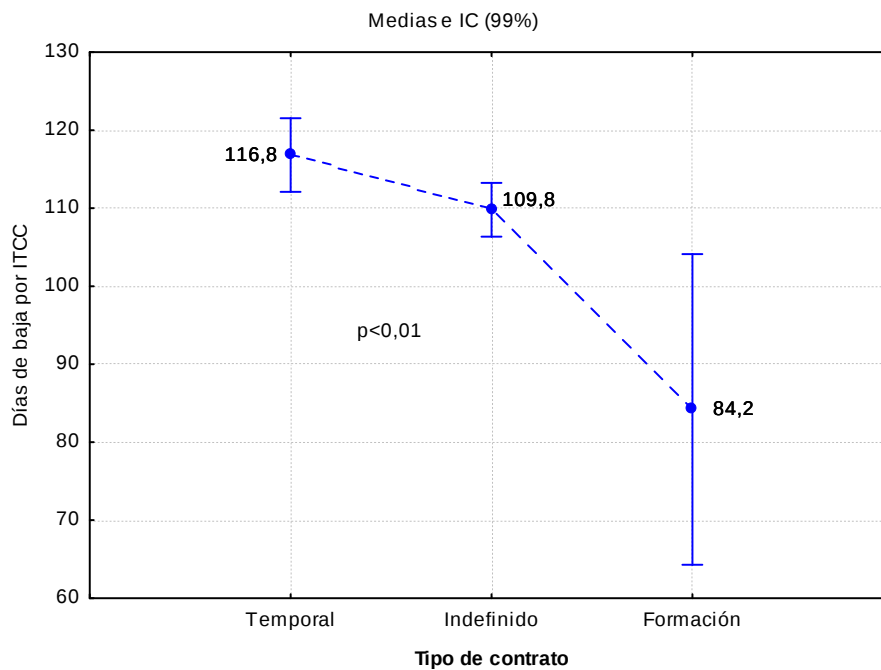
Figura 102. Duraciones medias de las bajas por cada uno de los sectores económicos.



Duración de la ITCC y tipo de contrato

Hay diferencias estadísticamente significativas entre los 3 tipos de contrato ($p < 0,01$) de modo que la baja tiende a durar más en los individuos que tienen un contrato temporal que en aquellos que lo tienen indefinido; si bien la diferencia no es muy grande (unos 7 días de media). Las diferencias son mucho más acusadas para el contrato en formación, aunque en este caso la muestra es mucho más pequeña (Figura 103).

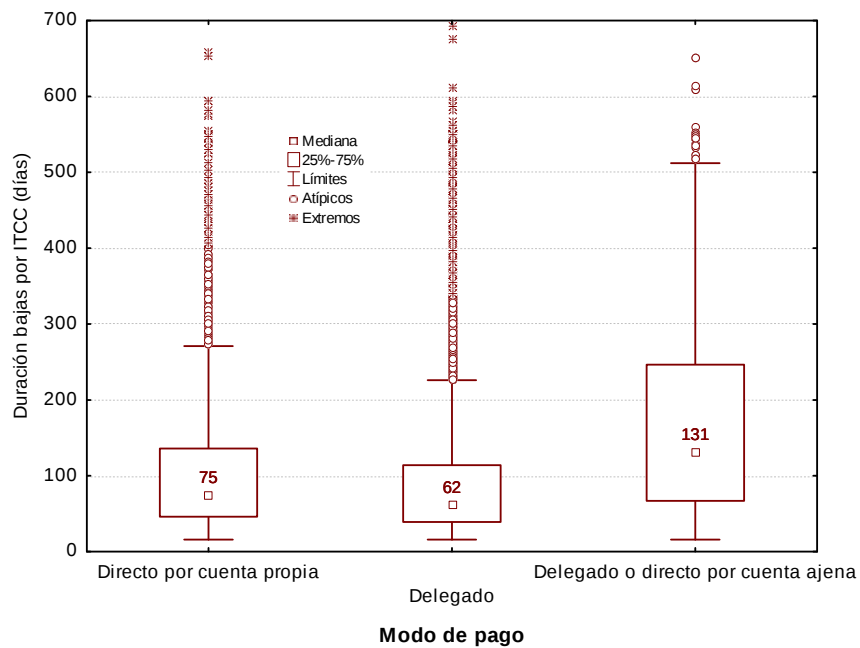
Figura 103. Duraciones medias de las bajas por tipo de contrato.



Duración ITCC y modo de pago

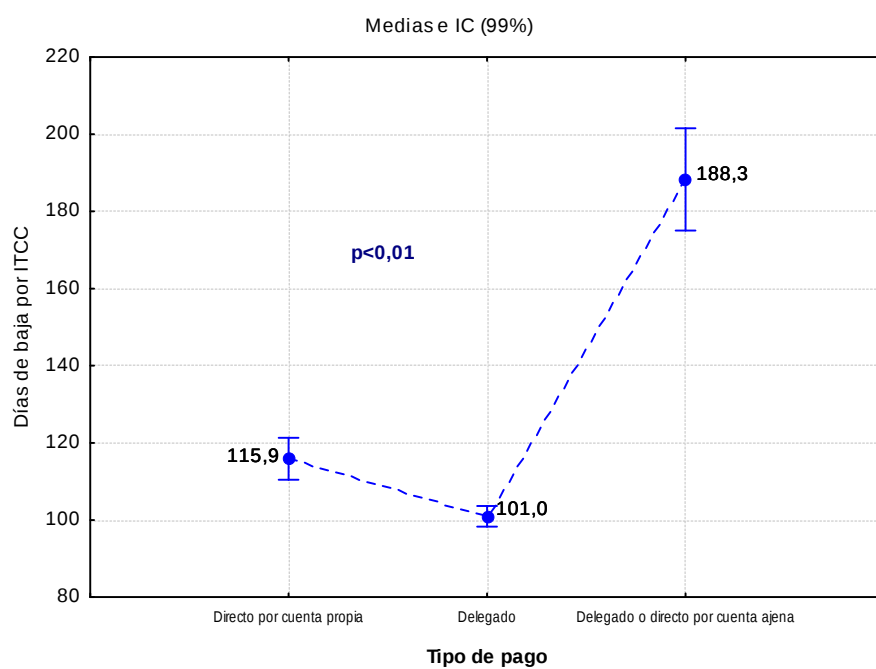
Las duraciones tienden a ser más cortas para el modo de pago “Delegado” (mediana igual a 62 días); es decir, para aquellos trabajadores que continúan vinculados contractualmente con la empresa durante el periodo que dura la baja. Las duraciones más largas, y con mayor dispersión de los valores, se corresponden con el modo “Delegado/directo por cuenta ajena” (mediana igual a 131 días); o sea, para aquellos trabajadores cuya vinculación contractual con la empresa se extingue durante el periodo que dura su baja laboral (por cualquier causa) o cuya empresa no puede hacer frente al pago del coste de su IT. El tercer modo de pago, “Directo por cuenta propia”, presenta una distribución con un perfil y unas estimaciones de los parámetros intermedios entre los dos primeros (mediana igual a 75 días). Este modo corresponde a los trabajadores autónomos. En los modos de pago “Delegado” y “Directo por cuenta propia” las bajas raramente (y son casos atípicos o extremos) superan los 230 días, mientras que en el modo “Delegado/directo por cuenta ajena” puede llegar a los 516 días e incluso superarlos (valores atípicos). Ver Figura 104.

Figura 104. Duraciones de las bajas laborales por modo de pago.



Encontramos diferencias estadísticamente significativas ($p < 0,01$) entre todas las categorías de *Modo de pago*. La duración media de la baja es claramente mayor para el pago “Delegado/directo por cuenta ajena” (188 días frente a 116 para el pago “directo por cuenta propia” y 101 para el pago “Delegado”). Hay menores diferencias entre el modo de pago “Delegado” y el “Directo por cuenta propia”, aunque la duración media de la IT sea mayor para éste último (unos 15 días). Ver Figura 105.

Figura 105. Duraciones medias de las bajas laborales por modo de pago.



Duración de la ITCC y pluriempleo

No encontramos diferencias estadísticamente significativas ($p=0,66$) entre trabajadores pluriempleados y no pluriempleados. Además solo hay en nuestra muestra 272 trabajadores pluriempleados, es decir, el 1,9% de la muestra.

La ITCC y el medio de transporte utilizado

Encontramos diferencias estadísticamente significativas ($p<0,01$), excepto para la categoría “NS/NC”, entre todas las categorías de esta variable (Figura 106). Tampoco hay diferencias estadísticamente significativas entre las categorías “Desconocido” (es decir, dato perdido) y “Transporte público”. Es llamativa la diferencia entre los trabajadores que acuden al centro de trabajo en transporte público y los que acuden en coche (especialmente cuando conduce el propio trabajador). Posiblemente, no podamos ver en ello una causa de alargamiento de la IT, sino más bien una variable de confusión o una variable que sirva como marcador de pacientes que tienden a tener bajas más largas.

Figura 106. Duraciones medias por tipo de transporte.

