



Universidad de Valladolid

FACULTAD DE CIENCIAS SOCIALES, JURÍDICAS Y DE LA
COMUNICACIÓN

Grado en Administración y Dirección de Empresas

TRABAJO DE FIN DE GRADO

**Análisis de la duración de las bajas laborales en España en
2017**

Presentado por Óscar Bernardos Yagüe

Tutelado por Alfonso Moral de Blas

Segovia, 10 de Junio de 2019

Índice

Resumen – Abstract.....	3
Introducción y justificación del trabajo	4
Precedentes y revisión de literatura anterior	5
Capítulo 1	7
1. Metodología y adecuación de los datos	7
1.1. Tratamiento de los datos.....	8
1.1.1. Adecuación de la base de datos.....	8
1.1.2. Adecuación de las variables	9
1.2. Metodología.....	11
1.2.1. Estadística descriptiva y recorrido interdecílico de los tipos de lesión.....	11
1.2.2. Doble regresión de Mínimos Cuadrados Ordinarios.....	11
1.2.3. Análisis de Robustez.....	15
Capítulo 2	17
2. Resultados y conclusiones.....	17
2.1. Estudio descriptivo	18
2.2. Presentación de los Resultados.....	23
2.2.1. Doble Regresión. Primera parte	23
2.2.2. Doble Regresión. Segunda parte.....	26
2.3. Análisis complementario y resultados de interés	30
2.4. Conclusiones.....	32
3. Bibliografía consultada	37
4. Anexos.....	¡Error! Marcador no definido.
4.1. Anexos capítulo 1	¡Error! Marcador no definido.
4.2. Anexos capítulo 2	¡Error! Marcador no definido.

Resumen – Abstract

El presente trabajo busca analizar de una manera detallada y desde diferentes puntos de vista y alternativas, las duraciones de las bajas por accidente de trabajo en España para el año 2017 a partir de la Estadística de Accidentes de Trabajo que elabora el Ministerio de Trabajo y Seguridad Social. El trabajo se inicia con un análisis descriptivo de aquellos grupos de lesión considerados más importantes. Posteriormente se hace una regresión primero con las variables de un contenido más fisiológico y se añaden en una segunda regresión aquellas con una influencia más accesoria y que están relacionadas con el concepto de “riesgo moral”. De manera complementaria, se cualifica la clasificación inicial incrementando el número de lesiones hasta 629 tipos de lesión para ajustar más la estimación y evaluar como pasan a comportarse las variables. Por último y para una mejor comprensión, se hace un análisis de robustez para de forma más detallada analizar posibles interacciones interesantes, reflejando cambios importantes.

The present work has the aim of analyse in a detailed way and with different points of view and alternatives, the durations of sick leaves due to work accidents in Spain. It is made for the year 2017 from the Workplace Accidents Statistic of the “Ministerio de Trabajo y Seguridad Social”. It starts with a descriptive analysis of those injury groups considered more important. Subsequently, a first regression is done with the variables of a more physiological content, and after in the second regression are added those with a more accessory influence (and related with the concept of “moral hazard”). In a complementary way, the initial classification is qualified with a higher number of injuries up to 629 groups to do a better adjustment and evaluate the performance of the variables after the change. Finally, and for a better comprehension, an analysis of robustness is done to analyse in a more complete way possible interesting interactions, reflecting important changes.

Introducción y justificación del trabajo

El presente trabajo analiza en detalle un fenómeno muy importante desde el punto de vista económico como es la duración de las bajas laborales, así como los factores que influyen en ella. Para poder realizar dicho estudio, se ha contado con la base de datos de los accidentes de trabajo en España de 2017. Se conoce como la Estadística de Accidentes de trabajo (ATR) que elabora el Ministerio de Trabajo, Migraciones y Seguridad Social. A partir de ella, se desarrollarán varias técnicas estadísticas y econométricas con las cuales se pueda analizar estos elementos y sacar conclusiones. La base de datos dispone de una cantidad suficiente de registros, y a su vez dispone de un amplio número de variables que capacitan para poder ajustar las regresiones con una gran significatividad.

De forma principal, se formula como hipótesis o principal objetivo el analizar los factores más determinantes de la duración de una baja laboral, a partir de elementos a priori que consideraremos como relacionados directos con el accidente. Y otros elementos que, aunque teóricamente no deberían influir en la baja, estudiaremos en búsqueda de explicar las diferencias entre duraciones para dos individuos estadísticamente idénticos en los factores inherentes a la baja. Todo ello poniendo atención a las diferentes amplitudes de la duración que se puedan observar a partir de los tipos de lesión. El incentivo de este trabajo es descubrir cuanto puede llegar a variar una baja por un tipo de lesión en concreto de un individuo a otro y los valores entre los que se pueden mover cada tipo de lesión. Es decir, para dos individuos con el mismo tipo de lesión y características propias de la lesión (grado, recaída o no, etc.) ver si puede darse el caso de que el tiempo de la baja sea mayor en un individuo que en otro, y en caso afirmativo ver en qué medidas y qué variables son las que influyen en este hecho.

Para ello se ha profundizado mucho en el estudio de todos esos posibles factores teniendo en cuenta estudios anteriores que aportasen argumentos que sirviesen para encaminar el estudio y a la vez tomando ideas de los posibles métodos para estudiar este fenómeno teniendo en cuenta después los que mejor podrían adecuarse a este caso. Se da particular importancia a los grupos de tipos de lesión, pues se estudia muy en detalle la distribución de los períodos de duración dentro de cada grupo y se focaliza el análisis en comparar unos con otros y ver cómo influyen los factores anteriores en unos u en otros.

Se trata de un tema influyente en el entorno empresarial y de las mutuas de seguros de accidentes de trabajo. Según la Seguridad Social se considera como contingencias profesionales a “los sucesos ocurridos durante el desarrollo de una actividad laboral y que originan alteraciones de la salud que se consideran accidente de trabajo (AT)”. Más en concreto la definición legal de accidente de trabajo y las condiciones establecidas figuran en el artículo 156 del texto refundido de la Ley General de la Seguridad Social, aprobado por el Real Decreto Legislativo 8/2015, de 30 de octubre para los trabajadores por cuenta ajena (en el caso de trabajadores por cuenta propia esta información se recoge en los artículos 316 y 317 de dicha ley). Este fenómeno al darse durante la actividad laboral y suponer una incapacidad temporal para trabajar, se muestra como trascendental para las empresas ya que ese trabajador no podrá volver a su actividad en un período de tiempo dependiendo de diversos factores, que serán los que analizaremos aquí. Por tanto, este análisis puede resultar muy útil para el mundo de las empresas y en particular a uno de los agentes que interactúan con el fenómeno de manera directa como son las mutuas de seguros por accidentes de trabajo.

En los próximos epígrafes se presenta, en primer lugar, la literatura previa existente relacionada con el estudio. En el primer capítulo se explicará como se ha adecuado la

base de datos y las variables para su procesamiento, y seguidamente, la metodología aplicada para las diferentes técnicas utilizadas. El capítulo de resultados y conclusiones muestra en primer lugar un análisis descriptivo de la duración de las bajas laborales para los diferentes grupos de lesión, luego presenta los resultados de las dos regresiones principales realizadas poniéndolas en confronto. Para acabar este capítulo se comenta la regresión con los grupos de lesión agrupados según el Cie-9 y las principales interacciones que se obtienen examinando las relaciones entre ciertas variables. Se termina el estudio con una breve conclusión y una lista con la bibliografía utilizada.

Precedentes y revisión de literatura anterior

En relación con este campo de análisis correspondiente a la relación entre economía laboral y economía de la salud, encontramos diversos trabajos que tratan el tema desde diferentes perspectivas. Primero de un modo más general y amplio se localiza un trabajo que presenta en sus anexos una tabla que muestra para distintas fechas, países, y temas particulares una gran recopilación de trabajos anteriores relacionados con el absentismo y la baja laboral (Osterkamp y Röhn ,2005). En el vemos que existen bastantes estudios que comparan este fenómeno entre países, otros que tratan individualmente un país, y en general vemos que es un hecho estudiado en muchos países tanto europeos como de América del Norte.

Por un lado, existe una línea de trabajos que se han centrado en la compensación del trabajador. Esta la recibe el trabajador si está de baja laboral y por ello se ha buscado ver como las variaciones de esta compensación, influyen en el número de días que dura la baja laboral. Véase para ello Dionne y St-Michel (1991). Un trabajo posterior (Bolduc, Fortin, Labrecque y Lanoie, 2002) amplía este estudio buscando no solo mostrar la influencia que esta compensación del trabajador tiene en los días de baja laboral, sino en la composición de las lesiones contabilizadas. Particularmente sobre este tema para el caso español se encuentra el trabajo de Martín-Román y Moral (2016) o los mismos autores, pero más antiguo y con otro enfoque (Martín-Román y Moral, 2005)

Ambas investigaciones, aunque se centran en la compensación al trabajador, muestran dos argumentos en los que este trabajo se apoya para realizar el análisis. Por una parte, más general, fija que los trabajadores se ven influenciados por incentivos económicos a la hora de darse de baja laboral e inciden en la duración de esta. En este caso los incentivos económicos serían la compensación del trabajador, y para este trabajo, no es el único factor que tomamos como incentivo económico, pues también se toman otros factores económicos considerando que pueden influir. Por el otro lado, sostienen el argumento de que esta variabilidad en el período de baja es mayor o menor de unos accidentes a otros, y que, a su vez, los cambios ante incentivos son mayores para unos accidentes que para otros pues existen diferencias entre las facilidades de diagnóstico, variabilidad o conocimiento de cada tipo de lesión. Es decir, dan peso a la hipótesis que se va a evaluar sobre si las bajas laborales tienen un componente fisiológico y uno más personal o moral para cada individuo, y que estos componentes no se reparten igual en cada tipo de lesión. Existe a su vez un trabajo (Markussen, 2012) que dando como base que la indemnización influye en el comportamiento, estudia el coste para cada individuo de estar de baja laboral.

Una asunción que podría surgir intuitivamente es que si un año aumentan las bajas laborales es debido a que se han empeorado las condiciones de trabajo. Es decir, sería normal considerar las condiciones del lugar de trabajo y los riesgos laborales como un

factor clave a la hora de determinar una baja o su gravedad. A este respecto, dentro de otra de las líneas en las que se ha investigado sobre la relación entre la economía laboral y de la salud, la ciencia empírica no dice lo mismo (Boone y Van Ours, 2002). Todo ello por tanto refuerza el objetivo del análisis en el cual se busca ver como el individuo ante los incentivos puede influenciar en el intervalo de baja laboral o incluso la baja en sí.

Más parecido al objeto de esta investigación, aunque aplicado a las contingencias comunes en vez de a las profesionales, existe un trabajo de Albertí, Jardí, López, Rocamora, Delclós y Benavides (2012), en el cual se estudia en detalle el período de baja junto con el número de casos totales y de cada grupo de lesión.

Por último, más cercano es el trabajo de Martín-Román y Moral (2017), pues desarrolla una investigación la cual da la base de este análisis ya que formula la idea de separar en dos componentes explicativos a la baja laboral. No obstante, el objetivo es diferente y con distintos métodos pues los autores se ciñen a la definición de absentismo para cuantificar el coste del absentismo en España (para EEUU también hay análisis parecidos como el de Waehrer, Dong, Miller, Haile y Men (2007), que cuantificaba el coste de las bajas laborales para la industria de la construcción) y buscan para ello utilizar solo la componente de riesgo moral. Para ello usan un método llamado “fronteras estocásticas” que ya previamente se había usado para un tema similar en Martín-Román y Moral (2014) separando trabajadores españoles y extranjeros, pero que en el estudio más reciente es más cercano aún a esa base utilizada aquí. En esta visión solo el riesgo moral produce estrictamente el sobre coste asumido dado que la baja en sí tiene un período mínimo asociado de recuperación que el individuo no puede evitar. En el caso de este análisis no se centra tanto en cuantificar el absentismo laboral sino en los grupos de lesiones y por ello se usa aquí una desagregación mayor de los tipos de lesiones.

Vemos por tanto que este trabajo es novedoso en el sentido de que este análisis no ha venido siendo muy común, pues es un tema que ha sido poco tratado desde el enfoque que aquí se le busca dar. Concretamente para el caso de España aún no existe mucha literatura al respecto, y al centrarse solo en un año se profundiza mucho en todos los detalles y variables de manera que se obtiene un análisis muy completo. Además, se apoya en anteriores investigaciones y estudios, que apoyan los argumentos de partida, con lo que a priori de realizarse el estudio aumenta las posibilidades de saber que se encontrará un resultado acorde con la hipótesis que previamente se plantea, pero sin ser información ya sabida, pues no se ha planteado de esta manera.



Universidad de Valladolid

Capítulo 1

1. Metodología y adecuación de los datos

1.1. Tratamiento de los datos

1.1.1. Adecuación de la base de datos

Como ya se ha mencionado previamente, la base de datos utilizada ha sido la Estadística de Accidentes de Trabajo, pero se aclara más en detalle ahora, que se ha trabajado con un fichero de micro datos con toda la información anonimizada en relación a Accidentes de Trabajo con baja, donde se recogen los registros de accidentes y recaída para los distintos períodos. Primero de todo como consideración previa se aclara que como viene explicado en el informe metodológico de la ATR “los accidentes de trabajo cubiertos por esta estadística son los sufridos por los trabajadores afiliados a la Seguridad Social con la contingencia de accidentes de trabajo específicamente cubierta”. Por tanto, no se incluyen aquellos sin la cobertura específica ni los “accidentes de trabajo sufridos por funcionarios mutualistas cubiertos por MUFACE, ISFAS y MUGEJU” (esto es importante pues son mutuas de trabajadores funcionarios, por lo que habrá que tener en cuenta la existencia de bajas laborales que no estén incluidas en esta base de datos). Tampoco están cubiertas las bajas de las ocupaciones militares.

Para el estudio de este fenómeno se ha utilizado el paquete estadístico “E-Views 10” a fin de poder tratar todos los datos de una manera adecuada. Previamente a procesar estos datos en dicho software se han eliminado ciertos registros que hemos considerado que podían afectar a la estimación.

En primer lugar, se han excluido los datos relativos a muertes, ya fuesen estas registradas en el campo “Grado de lesión” o “Grado de lesión en el momento del alta”. Esto se hace debido a que los accidentes laborales que causan una muerte o cuyo desenlace es una muerte afectan a la variable duración que nosotros queremos estimar, ya que no por ser el nivel de gravedad de lesión más alto suponen un mayor tiempo de baja. Lo cual conduce a eliminar estos registros buscando así centrarnos mejor en el fenómeno particular a investigar.

A su vez, en este paso, se ha detectado que en el campo “Grado de lesión en el momento del alta” existían datos inexistentes o vacíos, y que de la misma manera esas filas tampoco presentaban datos en el campo “Fecha del alta médica”, por tanto, ante la posibilidad de que por causa de esto la duración de la baja lleve consigo errores, se han omitido también estos registros dado que cabía la posibilidad de que afectasen a nuestra variable principal. En el archivo acerca del diseño de registro especifica que “Si la fecha de alta es desconocida este número de días se imputa estadísticamente”, por lo tanto, respecto a esos datos vacíos se concluye que lo mejor es eliminarles pues sino estarían sesgando el análisis.

En segundo lugar, en el campo “Tipos de contrato” se detectan datos donde figura un 0. En el diseño de registro no figura que este valor corresponda a ningún código que se refiera a un tipo de contrato concreto por lo que se eliminan también estos registros debido a que esta variable será usada posteriormente y estaría afectando a nuestra estimación. Por último y residualmente de la misma manera se detectan ceros en el campo “Nacionalidad” y al corresponder a “no consta” se eliminan pues también es una variable que se usará.

Por tanto, se ha hecho una adaptación de los datos que había de inicio para evitar posibles errores o datos que hayan sido mal recogidos (errores de dedo o registro) y por tanto afecten a la calidad de la estimación. De esta manera, para el período de 2017 se partía

aproximadamente con un poco más de 624.000 observaciones y debido a las correcciones el número de observaciones se ha quedado en 581.000 aproximadamente. No obstante, sigue siendo un número suficiente como para poder realizar la estimación, y facilitará el ajuste al haber eliminado posibles errores de registro.

1.1.2. Adecuación de las variables

Una vez tratados los datos y adaptados para poder hacer mejor la estimación, se han procesado con el software mencionado. Previo a la realización de las regresiones se han tratado ciertas variables para que luego puedan ser añadidas en la regresión. En este trabajo, y con el objetivo de no extenderse más de lo necesario en el apartado de metodología, se explica a continuación brevemente el tratamiento que se ha dado a aquellas que hemos utilizado y consideramos importantes de explicar para su correcta comprensión.

Simplificando, primero se han transformado varias variables cuyos registros eran “1” y “2” en variables dummy “0”, “1”. Asimismo, otras variables se ha considerado dejarlas como variables continuas. Posteriormente, las variables que presentan más de dos posibles registros en ellas, también se han transformado en variables dummy, pero presentan tantos dummy para dicho campo como posibles registros o grupos de registros tiene. De estas últimas, se explican en detalle a continuación las que son clave debido a que las introducimos en alguna o varias ecuaciones y en las cuales nos fijaremos después para el análisis.

La variable tipo de lesión o descripción de la lesión es fundamental en nuestro análisis y está clasificada siguiendo los criterios: “Codificación según Orden TAS/2926/2002” en donde hallamos grupos de tipo de lesión y a su vez subgrupos. En diferencia con otros trabajos y aprovechando la gran cantidad de datos que se tienen disponibles, se ha intentado agregar lo menos posible los tipos de lesiones, con el fin de poder obtener una mejor visión a fondo y así poder ver claramente donde están los datos que nos puedan aportar más información o aquellos que se comporten de una manera muy diferente al resto.

Por ello, hemos mantenido la clasificación atendiendo a los subgrupos, salvo en aquellos grupos que no se subdividían. Previamente se ha buscado en literatura anterior como se clasificaba esta variable. Se observan trabajos como el de Comisiones Obreras (García, Gadea, y López, 2007) en el cual dentro de todo el estudio que realizan, exponen tablas de diferentes análisis donde figuran diferentes formas de clasificaciones de los accidentes laborales. Ya sea por grupos principales de enfermedades o por categorías según el CIE-9 o el CIE-10 etc. En referencia a esto y profundizando un poco más la Clasificación Internacional de las Enfermedades se planteaba como la mejor alternativa a categorizar los tipos de lesiones, pero los datos que muestra la encuesta no son homogéneos ni muestran posibilidad de ser transformados de manera que se pueda hacer una clasificación acorde. Por ello en este trabajo se considera como mejor opción usar la propia clasificación que viene, pero de la manera más desagregada posible y así en este trabajo, se obtienen 37 tipos de lesión. Además, se debe tener en consideración que la codificación se basa en una recomendación de la OIT (Organización Internacional del Trabajo) y Eurostat, lo que nos da una mayor fiabilidad. Aun así, debido a que la CIE-9 figura entre las variables registradas en la ATR, se usará para ver en detalle lesiones o enfermedades muy concretas (sin agrupar).

También dentro de cada uno de los códigos se ha comprobado que existiesen datos en todos ellos y se ha analizado la distribución de esta variable para asegurar la ausencia de

Capítulo 1

errores o desviaciones. Estos 37 tipos de lesión que vamos a utilizar para estudiar el fenómeno son los que figuran en la Tabla 1.1 (en los anexos de este capítulo), donde se puede observar una breve descripción asociada a cada código asignado. Se aclara que el primer conjunto de datos que figura en el diseño de registro es “Lesión desconocida”, pero que ha sido colocada con el número 2, pues se ha preferido utilizar “Lesiones superficiales y cuerpos extraños en los ojos” como número 1, para que esta fuese la que no figure en la ecuación y por tanto venga incluida en la constante. Con ello se busca facilitar las comparaciones una vez obtenidos los resultados, al tratarse de una variable más concreta que “Lesiones desconocidas”.

Por otro lado, la variable partes del cuerpo atiende también a la “Codificación según Orden TAS/2926/2002”, la cual divide estas en grupos y subgrupos. En este caso se han escogido agregaciones de los grupos como se muestra en la Tabla 1.2 que se encuentra en los Anexos del capítulo 1.

También se ha agrupado la variable edad. Para el criterio de las agrupaciones se ha tenido un poco en cuenta las diferentes etapas, las diferencias que puede haber en cómo puede afectar una lesión según la edad y además la distribución de las edades para no dejar grupos con pocos datos. Con unos intervalos no muy amplios y teniendo en cuenta las etapas que pueden tener los individuos en su vida laboral y de salud se agrupan de la siguiente manera presentada en la Tabla 1.3 (que figura en los anexos del primer capítulo).

En el caso de la variable “Grado de alta”, que se refiere al Grado de lesión en el momento del alta, es más simple, pues se tenían solo 3 grupos donde el primero incluía a las lesiones de grado leve, el segundo a las de grado grave, y por último en el grupo 3 están las muy graves. En nuestro análisis para simplificar y dado que evaluando la distribución de los datos en esta variable se han observado pocos datos en el grupo “Muy grave”, se separan solo en dos grupos las variables. Uno con las lesiones leves llamado gravedad baja, y otro con las graves y muy graves, que se llamará de forma genérica, grado grave.

Para la segunda ecuación también se han tratado ciertas variables como por ejemplo la variable sobre el tipo de contrato que tenía el trabajador en el momento del accidente. Tipo de contrato es un campo que les clasifica por indefinidos, temporales u otros y a su vez por tiempo completo o parcial. Sin embargo, para este trabajo, se considera que de ambas características la que más puede afectar en este caso es la seguridad de tener un trabajo por tiempo indefinido frente a saber que lo tienes con una fecha de caducidad (ya que el tener una fecha fijada de finalización puede motivar a que no se busque regresar al trabajo con la misma prisa que sin fecha fijada, salvo en el caso de que existiera la posibilidad de pasar de temporal a indefinido). Por ello se van a clasificar solo en 3 grupos: Indefinidos (Del código 100 al 399), temporales (del 400 al 600) y otros contratos (del 600 en adelante).

En el caso de la variable “Situación profesional” esta simplemente se ha formado con ella 4 variables dummy donde la primera hace referencia a “Asalariados del sector privado”, la segunda “del sector público” y la tercera y cuarta autónomos con y sin asalariados a su cargo respectivamente. En esta ya en estudios previos se presentaba como diferente el grupo perteneciente a autónomos, visto en el análisis de Martín-Román, Moral y Corales-Herrero (2013). Otra agrupación realizada es para los sectores económicos. Para ello con los datos que se tienen en función de la Clasificación Nacional de las Actividades Económicas (CNAE-2009) se han sacado 5 grupos: sector agrario del código 011 al 032 (agricultura, ganadería, silvicultura y pesca), sector de industrias extractivas del código 051 al 099, sector industrial de los códigos 100 al 400 (industria manufacturera y suministros), sector de la construcción del 410 al 440 y por último el sector servicios de

los códigos 450 hasta 999. Relacionado con esto se han agrupado también los tipos de ocupación en dos grandes grupos utilizando los códigos relativos a la Clasificación Nacional de las Ocupaciones (CNO-2011). El criterio de agrupación es la predominancia o no de actividades físicas en la ocupación, obteniendo de esta manera un grupo donde no predominan las actividades físicas que va de los códigos 100 hasta 500, y el otro grupo que va de los códigos 500 hasta 999.

Otras variables que han sido preparadas para su utilización serían la Nacionalidad, el día de la semana en el que se produce el accidente y la estación del año todas ellas transformadas simplemente en dummies. También la variable lumbago ha recibido el mismo tratamiento. En este caso usando la codificación CIE-9-MC, donde lumbalgia corresponde al código 724.2 de la clasificación.

1.2. Metodología

1.2.1. Estadística descriptiva y recorrido interdecílico de los tipos de lesión.

Previo a realizar este primer análisis, en busca de complementarlo y facilitar una mejor comprensión, se han estudiado las estadísticas descriptivas de la duración de la baja laboral para cada tipo de lesión. Se hace en concreto para cada tipo de lesión debido a que es la variable principal que se está analizando y que por ello es más relevante para este trabajo. Con ello, en esta parte del análisis se pretende centrar la atención sobre las distribuciones de las duraciones en cada tipo de lesión, y ver de esta manera la amplitud/asimetría y forma de cada una de ellas para comparar entre los distintos tipos de lesión. Así se verá de manera bastante visual como hay unas variables que son más difíciles de estimar de una manera ajustada debido a que tienen una alta variabilidad y que su distribución no se ajusta a una distribución con un coeficiente de curtosis igual o cercano a cero. En otros términos, podremos ver si los datos están muy concentrados hacia la media o menos concentrados (Distribución leptocúrtica o platicúrtica).

Dentro de la información descriptiva que se sacará además de la media están la moda, la mediana, el coeficiente de asimetría y el coeficiente de curtosis. Pero en el que este análisis se va a centrar es en el rango interdecil, que se trata del rango entre el percentil (X,10) y el percentil (X,90), dado que se considera que puede mostrar muy claramente entre que rangos se encuentran el 80% de los valores del centro, pudiendo así comparar de una manera muy sencilla los rangos de duración de unos tipos de lesión con otros. Esto ya se ha realizado en varios trabajos anteriores como se puede observar en la investigación de “Catalina-Romero, Martínez-Muñoz, Quevedo-Aguado, Ruiz-Moraga, Fernández-Labandera, y Calvo-Bonacho (2013) donde se utiliza el recorrido intercuartílico para mostrar esa variabilidad de cierto tipo de enfermedad y además poder comparar entre las que analiza.

Este proceso se hará para los tipos de lesión que se considera que pueden ser relevantes, ya sea porque se espere una mayor o menor amplitud y se busque comprobarlo, o porque se crea que tienen una importancia o peso relativo alto.

1.2.2. Doble regresión de Mínimos Cuadrados Ordinarios.

Para comenzar se va a realizar una estimación de la duración mediante una regresión logarítmica de mínimos cuadrados. Como ya hemos visto de trabajos anteriores, las contingencias, en este caso las profesionales, han sido muy estudiadas con ánimo de

esclarecer el componente explicativo que llegan a alcanzar diversos factores relativos al accidente. En parte también se ha buscado constantemente obtener las duraciones medias para cada accidente laboral teniendo en cuenta ciertos parámetros. De todas ellas podemos sacar en claro la hipótesis o el argumento de que el tiempo de baja laboral puede verse afectado por parámetros no relacionados directamente con el accidente en concreto. Uno de los trabajos que más variables planteó como esas variables indirectas que pueden influir en el período de baja laboral es el de Krueger (1990), donde se estudian los parámetros asociados directamente con la lesión, pero a su vez menciona otros posibles factores que se contrastan en hipótesis para comprobar la veracidad de su influencia o no en un caso empírico. Todos estos, aunque no existe un consenso de cuales son, se considerarían factores de riesgo moral. Pues como se puede comprobar diferentes variables son tomadas para el caso español también (Como por ejemplo en de Blas, Román y Caballero, 2013) que utilizan variables de región, tipo de contrato etc.

Este riesgo aplicado a nuestro fenómeno serían todos aquellos factores que conforman diferentes duraciones para dos individuos estadísticamente iguales en los parámetros directamente relacionados con la lesión. Es decir, podríamos hablar de parámetros subjetivos, que muestran diferencias en cada individuo, ya sea porque a ese individuo le afecte más un accidente que a otro, o por otros factores que rodean la posibilidad de que este influya en la duración. Por ejemplo, si se habla de dos individuos de género masculino, de misma edad, que han sufrido un accidente de la misma gravedad y clasificados con el mismo tipo de lesión y todas las variables relativas iguales, teóricamente ambos tendrían una misma duración de la baja. Como se observa que esto no ocurre, se habla de la posibilidad de que esa variabilidad se deba a otros factores particulares de cada individuo. Por ello separaremos dentro del tiempo de baja, una parte de él que se explicaría por factores fisiológicos asociados al accidente, y la otra parte se explicaría por otros factores que consideraremos como ya se viene haciendo, de riesgo moral. Así tendríamos que el período de incapacidad temporal de cada individuo tiene dos componentes, uno que llamaremos fisiológico, y otro moral o de riesgo moral.

En consecuencia, para el análisis se formulan dos ecuaciones en las que se incluirán primero solamente el grupo de variables que consideramos que están relacionadas de manera directa con el accidente laboral, y por otro lado una segunda regresión a la que se añadirán ese segundo componente en el que en este trabajo se divide a la duración, es decir, las variables de riesgo moral. Con ello, al aislar primero en una ecuación solo lo considerado como factores fisiológicos, y en una segunda todos, se podrá ver y analizar cómo cambia la capacidad explicativa del modelo y como cambia la estimación.

La primera ecuación por tanto consta de aquellos factores que se han considerado que tienen una relación directa con la baja laboral. Por un lado, como variable dependiente ya se sabe que está el logaritmo de la duración de la baja. Por tanto, al expresarlo como una función logarítmica se busca un mejor ajuste, y se deben tomar ciertas consideraciones en cuanto a la posterior transformación de los coeficientes para poder interpretarlos. La transformación es simple como podemos ver por ejemplo en un estudio de Velasco (2004), que usa la regresión logarítmica y explica de una forma clara como transforma luego esos coeficientes para su interpretación. No es más que despejar la siguiente fórmula: $(e^{\text{Coeficiente}} - 1) * 100$. Es decir, elevamos el número “e” al coeficiente de la variable que queremos transformar, le restamos uno, y al multiplicarlo por cien obtendremos el incremento porcentual que esta supone. Como ejemplo, un coeficiente de 0.1401, despejando significaría que esa variable supone un aumento del 15.04% con respecto a la variable o dummy que hayamos dejado fuera.

Metodología

Las variables independientes serían: las consideradas como dos principales variables demográficas que influyen en el fenómeno, que son el género y la edad. El género se añade aquí dado que existen ciertas enfermedades o lesiones que pueden tener un diferente desarrollo por diferencias fisiológicas, como pueden ser roturas de huesos, que pueden afectar de diferente manera. Ya para España, aunque no concretamente para el período de baja, hay un estudio acerca de las posibles diferencias de género del seguro por baja laboral hecho por el autor Martín-Román en 2008. Por su parte la edad aparece en varios grupos, por lo que podremos ver cómo afecta en mayor o menor medida a cada grupo de edad, y si crece la influencia según se sube en grupos de edad o decrece. Luego las siguientes variables son si la asistencia fue ambulatoria, si el accidente era una recaída, si hubo hospitalización y las tres últimas que hemos explicado en detalle su obtención anteriormente. Estas son el grado de lesión en el momento de alta, la parte del cuerpo lesionada, y el tipo de lesión según la descripción de esta. De esta manera la ecuación a estimar quedaría así:

$$\text{Log}(\text{Duracion}) = \alpha_0 + \vartheta_1 \text{ hombre} + \sum_{i=2}^8 \delta_i \text{ edad}_i + \theta_1 \text{ ambulatoria} + \theta_2 \text{ recaida} + \theta_3 \text{ hospitaliz} + \theta_4 \text{ gradograve} + \sum_{i=2}^9 \gamma_i \text{ parte}_i + \sum_{i=0}^{37} \beta_i \text{ tl}_i + \varepsilon$$

Ecuación 1.1. Variables fisiológicas. Elaboración propia.

Como se puede comprobar (Ecuación 1.1), las variables introducidas llevan relación directa con las características del accidente laboral, pues a priori, si uno se pregunta cuanto va a durar la baja laboral de un individuo, se fijaría en estos datos para así intentar calcular el período que este individuo podría estar de baja.

Con la formulación de la Ecuación 1 que se observa, se puede ver que por tanto la constante recogerá la duración media de un individuo mujer, con una edad comprendida entre 16 y 24 años, con una asistencia hospitalaria, con un accidente que no es recaída, sin hospitalización (en otras palabras, que el individuo no es ingresado), con un nivel de gravedad de la lesión leve, y que esta ha sido en la cabeza y se ha clasificado como “lesiones superficiales o cuerpos extraños en los ojos”.

En la segunda ecuación además de las variables anteriores, se añadirán aquellas que en este estudio se han considerado como variables de riesgo moral. Aunque teóricamente no deberían influir en el período de contingencia profesional, existe ese riesgo de que de unos individuos a otros existan diferencias debidas a factores más inherentes a cada individuo y a su situación personal.

Por ello como variables espaciales se añaden la nacionalidad presentada como una única variable que recoge si el individuo este nacionalizado español o no (“esp” corresponde a estar nacionalizado español). A priori vemos que no debería influir esto, pero sí que se considera de que exista esa posibilidad de que motive una diferencia y por tanto se añade. En términos de regiones, se incluye una variable para cada comunidad autónoma (y las dos ciudades autónomas, es decir en total 19 conjuntos, tomando para ello la variable de provincia de la autoridad laboral que recibe el parte de accidente de trabajo) que nos muestre y evidencie las diferencias o no entre que la baja laboral se haya producido en un lugar u otro de España. Se hace atendiendo a estudios previos que ya formulan y observan grandes diferencias entre las duraciones de las bajas laborales según la CCAA (Corrales, Román y Moral (2008)). En este sentido entran en juego factores acerca del funcionamiento del sistema de mutuas y médicos en cada CCAA de España en cuanto a la rigidez, rigurosidad y seriedad para dar una baja y cuantificarla entre unas y otras, a su vez la posible diferencia de mentalidad entre individuos de unas y otras Comunidades que

Capítulo 1

lleve a que en unas regiones se alargue más el período de Contingencia Profesional o no. También podrían influir las diferentes tasas de desempleo o salarios de las regiones e incluso podrían influir en los períodos de recuperación los distintos climas o condiciones meteorológicas de cada región (aunque este último sería un factor fisiológico, no de riesgo moral).

En segundo lugar y en referencia a motivos más económicos puros se han seleccionado varios factores. Se añade el tipo de contrato en esta parte, pues se piensa que este puede influir. Así pues, se espera que el contrato sea indefinido o no influya en que la duración sea distinta entre diferentes individuos. Es decir, se busca ver que efecto tiene la seguridad de saber que se tiene ese trabajo de manera indefinida frente a la seguridad de tenerle hasta cierta fecha. De la misma manera, se añade también la variable de cuantía diaria del subsidio por incapacidad temporal. Se añade como variable continua y con ello se quiere comprobar si el que un individuo reciba una mayor cuantía, puede hacer que el período de baja sea mayor o no. Por último, en este apartado también añadimos las variables relativas a la situación profesional con el fin de examinar si el trabajar para el sector público o para el privado o para uno mismo influye.

Luego en cuanto a factores más relacionados con el tipo de trabajo se suman dos variables. Por un lado, de los sectores de la economía al que pertenece cada individuo, donde se separan en los siguientes grupos: sector agrario, industrias extractivas, industria, construcción y sector servicios (La explicación de las agrupaciones se ha hecho anteriormente). En este último podrían influir los diferentes niveles de desempleo en cada sector o las condiciones de los sectores (en auge, en caída, estancados, mucha competencia...). Y por otro lado una variable que recoja por un lado aquellas ocupaciones donde predomine la actividad física o este componente esté más presente, y por otro lado aquellas donde este pase a un segundo plano. Con ello se quiere analizar si por el tipo de ocupación se busque evadir más o menos el trabajo o sea más importante una total recuperación de la lesión. Ejemplo de ello sería un esguince que para un trabajador de oficina no sería el mismo impedimento que para un albañil.

En los trabajos ya mencionados en la página 4 sobre la literatura anterior se añade como variable de riesgo moral una en la que se agrupan las lesiones fáciles de diagnosticar y otra con las difíciles de diagnosticar, atendiendo a diferentes criterios, pero donde llegan en común a señalar como regla general que los problemas de espalda sobre todo lumbagos y los esguinces son más difíciles de diagnosticar. A pesar de ello, no existe un claro consenso, y ante la falta de total conocimiento del tema, no se añade como en anteriores trabajos la variable de enfermedades de difícil diagnóstico. En su lugar se añaden estos dos grupos de lesiones mencionadas separadas cada una (en el caso de los esguinces no es necesario añadir ninguna dummy porque ya están dentro de los grupos t108, t109 y t110, aunque más en concreto en los dos últimos). Así se podrá analizar y centrar un poco más en detalle estas dos causas de baja laboral para comprobar si como la literatura precedente dice existe una mayor variabilidad en las duraciones y si empíricamente se observa mayores duraciones en estos grupos (es decir, sin tratarlas como de difícil diagnóstico pues no se sabe a ciencia cierta, sino que se examinan por separado para ver sus efectos). En el caso de los esguinces estos representan alrededor del 40% de las causas de bajas laborales y las lumbalgias alrededor de un 8%.

Por otro lado, también se añaden variables de tiempo, en la cual por un lado se incluye variables dummy de cada uno de los días de la semana, exceptuando uno que en este caso será el lunes, y de la misma manera cogiendo de 3 en 3 se añaden las estaciones del año, que aunque no se ajusten a las fechas exactas, nos permiten ver si existe una variación para los 3 primeros meses del año, los tres segundos meses etc. Como se ve, son variables que como ya se ha comentado a priori no deberían explicar el período de baja, pero para verificarlo se incluyen en el estudio para ver cómo se comportan en realidad. Con todo esto la expresión a estimar quedaría expresada de la siguiente manera:

$$\begin{aligned} \text{Log}(\text{Duracion}) = & \alpha_0 + \vartheta_1 \text{hombre} + \sum_{i=2}^8 \delta_i \text{edad}_i + \theta_1 \text{ambulatoria} + \\ & \theta_2 \text{recaida} + \theta_3 \text{hospitaliz} + \theta_4 \text{gradograve} + \sum_{i=2}^9 \gamma_i \text{parte}_i + \sum_{i=02}^{37} \beta_i \text{tl}_i + \\ & \theta_5 \text{esp} + \sum_{i=2}^{19} \rho_i \text{region}_i + \tau_1 \text{temporal} + \tau_2 \text{otroscontratos} + \theta_6 \text{cuantdia} + \\ & \sum_{i=2}^4 \varphi_i \text{sitprof}_i + \theta_7 \text{extractivas} + \theta_8 \text{industria} + \theta_9 \text{construccion} + \\ & \theta_{10} \text{servicios} + \theta_{11} \text{ocupacionnofisica} + \theta_{12} \text{lumbago} + \sum_{i=2}^7 \sigma_i \text{diasemana}_i + \\ & \sum_{i=2}^4 \omega_i \text{estacion}_i + \varepsilon \end{aligned}$$

Ecuación 1.2. Variables fisiológicas y de comportamiento. Elaboración propia.

Por tanto, en esta nueva regresión, la constante o “alpha sub-cero” recogería la duración media de un individuo con ciertas características. En concreto estas características serían las ya mencionadas para la primera ecuación por un lado y a mayores: Se trataría de un individuo no nacionalizado español, cuyo parte ha sido recibido en la Comunidad de Madrid, con un contrato indefinido, con una cantidad diaria de subsidio de la media de la población, que trabaja para el sector privado, en el sector agrícola en una ocupación donde predomina la actividad física, no es lumbago la causa de su baja laboral y el accidente se ha producido un lunes de los 3 primeros meses del año.

Como ya observamos la ecuación al final tiene una cantidad considerable de variables, por lo que se tendrá a disposición información para examinar muchos factores que no se sabe a priori si influyen en la duración de la baja o no, y en qué medida. Al haber hecho dos ecuaciones, podremos ver cómo cambian los indicadores de la regresión al añadir en el segundo paso las variables de riesgo moral, así como las propias variables que se han considerado como relacionadas directamente con la baja.

1.2.3. Análisis de Robustez

Por último, con ánimo de comprobar la robustez de las ecuaciones y estimaciones se ha hecho un análisis que permitiese ver que está bien hecha y ver si es posible ajustar más aún la estimación o no. Para ello se deben tomar algunas consideraciones, que posteriormente serán planteadas en el apartado de resultados. Con este objetivo, se ha hecho una comprobación de las principales interacciones entre variables. Así se pretende examinar las interrelaciones que puede existir entre las variables pero que no vemos a simple vista, y ver más en detalle cómo se podría mejorar el ajuste. Para ello se han usado las interacciones, que no es más que introducir las dos variables sobre las que se quiere analizar su interrelación multiplicándose. Al haberse realizado sólo con las variables dummy, simplemente con una variable que multiplique a ambas dos a analizar es suficiente. Al añadir estas interacciones en la ecuación podremos ver si algunas variables por si solas tienen un mayor efecto que el que pensábamos o un menor efecto. Para ello simplemente hay que fijarse en el coeficiente obtenido de esta variable. Este análisis es muy útil pues se ha realizado de una manera amplia, y después se presentarán aquellas

Capítulo 1

variables que se han considerado que presentan resultados relevantes, ya sea porque muestren un coeficiente en con signo contrario que separadas, o porque aumenten en mayor medida su efecto que otras. Por ejemplo, podríamos analizar las interacciones entre la variable hospitalización y la variable recaída, donde ambas presentan coeficientes positivos, pero que, al añadir su interacción, si esta presenta un coeficiente negativo mostraría que cuando ambas variables se presentan juntas, la duración crece menos que lo que creíamos a priori. Asimismo, sería relevante si presentara un coeficiente del mismo signo, pero alto, que nos mostraría que el efecto de ambas juntas es aún mayor del que se la ecuación nos mostraba.

Como anteriormente se ha dicho, investigando sobre la posible manera de como clasificar las lesiones se había encontrado la CIE-9 como la mejor opción pero que por ciertos problemas no es posible hacer grupos para utilizar en las ecuaciones principales. No obstante, buscando un mejor ajuste y profundizar todo lo posible en el análisis se va a realizar una regresión donde se eliminan los 37 grupos de tipo de lesión de la ecuación 2, y en su lugar se colocan como grupos los del CIE-9. Esta clasificación divide las enfermedades mediante códigos de 3 cifras enteras y 2 decimales. Así, se han utilizado las tres cifras enteras, con el decimal truncado (pues una enfermedad del código 739.9, posiblemente está más relacionada con el 739 que con el 740). Con todo este proceso en vez de 37 variables sobre tipo de lesión se han obtenido 1000 variables, con las cuales se podrá tratar de buscar que grupos son aquellos con una mayor significancia (medida por su estadístico-t en este caso) y una mayor relevancia, pudiendo así hablar de enfermedades de una manera muy concreta y al detalle. Por ello no se presentará la ecuación completa, sino que solo se citarán aquellos grupos de interés para el análisis.



Universidad de Valladolid

Capítulo 2

2. Resultados y conclusiones

2.1. Estudio descriptivo

Como inicio, se realiza este estudio previo, con el fin de conocer más profundamente las características de las distribuciones de los diferentes tipos de lesión, y que de esta manera se pueda comprender mejor y analizar con más criterio las regresiones posteriores. Se realiza acerca de las variables tipo de lesión, pues en concreto es uno de los puntos clave en los que se centra el trabajo, y más en particular, la atención se focaliza sobre los tipos de lesión más relevantes (ya sea por número de registros, o por interés para el análisis). Además, se estudia también en este apartado de manera exclusiva la variable “lumbago”, debido a la importancia ya mencionada, para así saber su distribución más a fondo.

Para comenzar, se comenta a continuación la distribución de frecuencias por días de la duración de la baja laboral para todos los individuos de España con cualquiera de los 37 grupos de lesión, es decir, incluyendo todas las lesiones. La baja laboral puede ser de hasta 545 días, y el número de casos va disminuyendo según va aumentando el tiempo, por lo que solo se muestra en el Gráfico 2.1 la distribución de frecuencias de las primeras 12 semanas. (A partir del día 85 en adelante, quedarían 51.545 datos aún, pero que en el total de la muestra es menos de un 10%, y haciéndolo de esta manera se puede ver más en detalle el principio que es dónde más registros hay concentrados)

A primera vista, se observa un hecho muy relevante, ya que se constata que en las duraciones 8, 15, 22...etc., es decir, en los múltiplos de 7 sumando una unidad ($7x+1$) existe un mayor número de registros claro. Este resultado va unido a un mayor número de registros los dos días anteriores. En términos generales, se aprecia una tendencia creciente hasta el octavo día, donde se presenta el mayor número de altas, con hasta casi 35.000 datos, luego lleva una tendencia descendente que va suavizándose. Si se excluye el día primero de baja que no tiene datos porque baja laboral se considera a partir del segundo día, se ve que contrariando a la tendencia general los dos últimos días de la medida semanal presentan menos datos, e inmediatamente después, el primer día de la siguiente semana que el individuo está de baja aparece con picos. Profundizando acerca de este fenómeno, se observa que todos los “primer día” de la semana son el 16.20% de los datos, los “segundo día” 15.85% de los datos, los “tercer día” 15.4%, los “cuarto día” 15.7%, los “quinto día” 15%, y luego los “sexto y séptimo día” son el 10.3 y 11.5 por ciento respectivamente.

Por tanto, se encuentra así un hecho con el que no se contaba. Aunque es muy difícil decir cómo influye a la duración, pues el hecho es que en un día de la semana hay de forma notoria más datos y en otros dos menos pero no se puede cuantificar cuánto afecta ello al fenómeno que aquí se estudia, que es el período de baja. A pesar de ello se postula como importante, y que en cierta medida puede estar sesgando la estimación, sin que se venga en ocurrencia una manera de adecuarlo o eliminarlo.

La explicación que se ve como hipótesis posible, es que la “medida” por defecto que usan los agentes que estipulan el tiempo de baja o el tiempo que tiene que estar un individuo hasta que vuelve para una revisión es la semana. Es decir, un individuo que sufre un accidente y le dan una baja para “dos semanas” parece razonable en cierta medida pensar que el individuo cumpla esas dos semanas de baja que le han sido prescritas. En muchos casos habrá necesitado esos 14 días, pero si se da el caso de que a los 12 días está bien, debería volver al trabajo pues la lesión ya está curada, pero esto no siempre podría suceder. Puede ser por precaución buscando una total recuperación, confianza en lo que le ha dicho un profesional, desconocimiento de que se está curado totalmente o también puede ser porque se tenga derecho a esas dos semanas y exista un riesgo moral de no

Resultados y conclusiones

volver dos días antes a pesar de estar ya bien. Es decir, existe una posibilidad (que se plantea como hipótesis, no puede confirmarse en absoluto) de que en ciertos casos aquí influya el riesgo moral, que será mencionado y examinado más adelante también, por lo que, aunque no se pueda cuantificar, se tiene en cuenta este fenómeno observado. Esta hipótesis explicaría que antes de cumplirse el período que marca la próxima revisión descendan las altas, y justo ese día aumenten mucho, ya que según la hipótesis los profesionales encargados de estipular esos tiempos de baja laboral, quizá por defecto usan la semana como medida de cuanto debe un individuo estar de baja o volver para ser revisado.

Se profundiza y se observa por separado distintos tipos de lesión con el objetivo de comprobar si el fenómeno se replica en todas las lesiones o no, o si se suaviza. Examinando se constata que existen diferencias, aunque no muy marcadas, con algunos tipos de lesión donde la distribución sigue esa tendencia, otras donde la tendencia es menos clara e irregular, y otras donde se puede ver que el fenómeno bastante bien. Por ejemplo, en “fracturas cerradas”, donde se disponen de poco más de 33.000 datos el primer día de la semana aparece más marcado. Por otro lado, en “esguinces y torceduras”, con más de 150.000 datos (y donde según la literatura es más probable que se den situaciones de riesgo moral), muestran mejor el hecho de que el sexto y séptimo día previos a la revisión semanal las altas registradas se reducen. Por tanto, aunque no en todas las lesiones se vea bien, ni en todas de la misma manera, se trata de un hecho bastante considerable para el estudio (En 2019 esta revisión ha dejado de ser semanal).

En cuanto a los parámetros estadísticos que describen a la distribución de todas las lesiones en conjunto, el período medio de baja es de 32.25 días, pero como ya se observa en el Gráfico 2.1, tiene una marcada asimetría positiva (hacia la izquierda), pues las altas se concentran en los primeros días, y a partir de la primera semana llevan una tendencia general decreciente del número de bajas. Esto se refleja numéricamente en que la mediana está en 15 y la moda 8, muy a la izquierda de la media, o en el alto coeficiente de curtosis que está alrededor del 18,5, o en el coeficiente de asimetría (CAF) que está en el 3,7, mostrando una distribución altamente sesgada hacia la izquierda y muy apuntada (mostrado por el coeficiente de curtosis), pero que concuerda con las características de la variable, ya que es una variable de duración y que por ello se puede equiparar a un modelo de supervivencia donde según se avanza en los días cada vez quedan menos “supervivientes” (personas de baja). La desviación estándar es muy alta con un 48,5 debido a la variedad que existe por coger todos los datos ahora. Para acabar, el decil 1 es el día 3, y el decil 9 el día 123, es decir, el 80% de los datos se concentran entre el día 3 y el 123.

Pese a que, para el estudio se ha hecho el análisis de las distribuciones de los 37 tipos de lesión (examinando sus distribuciones y sus principales parámetros estadísticos como son la media, mediana, moda, desviación estándar, asimetría etc.), aquí solo serán tratados aquellos tipos de lesión que son más importantes en el trabajo, ya sea porque son una parte representativa de la media o porque presentan resultados relevantes o determinantes para el estudio. Por ello solo se comentarán ciertos grupos de “tipo de lesión”. Abajo (Tabla 2.1) se muestra como ejemplo de la información disponible algunos de los principales grupos de lesión con los estadísticos a los que más importancia se han dado y comentado.

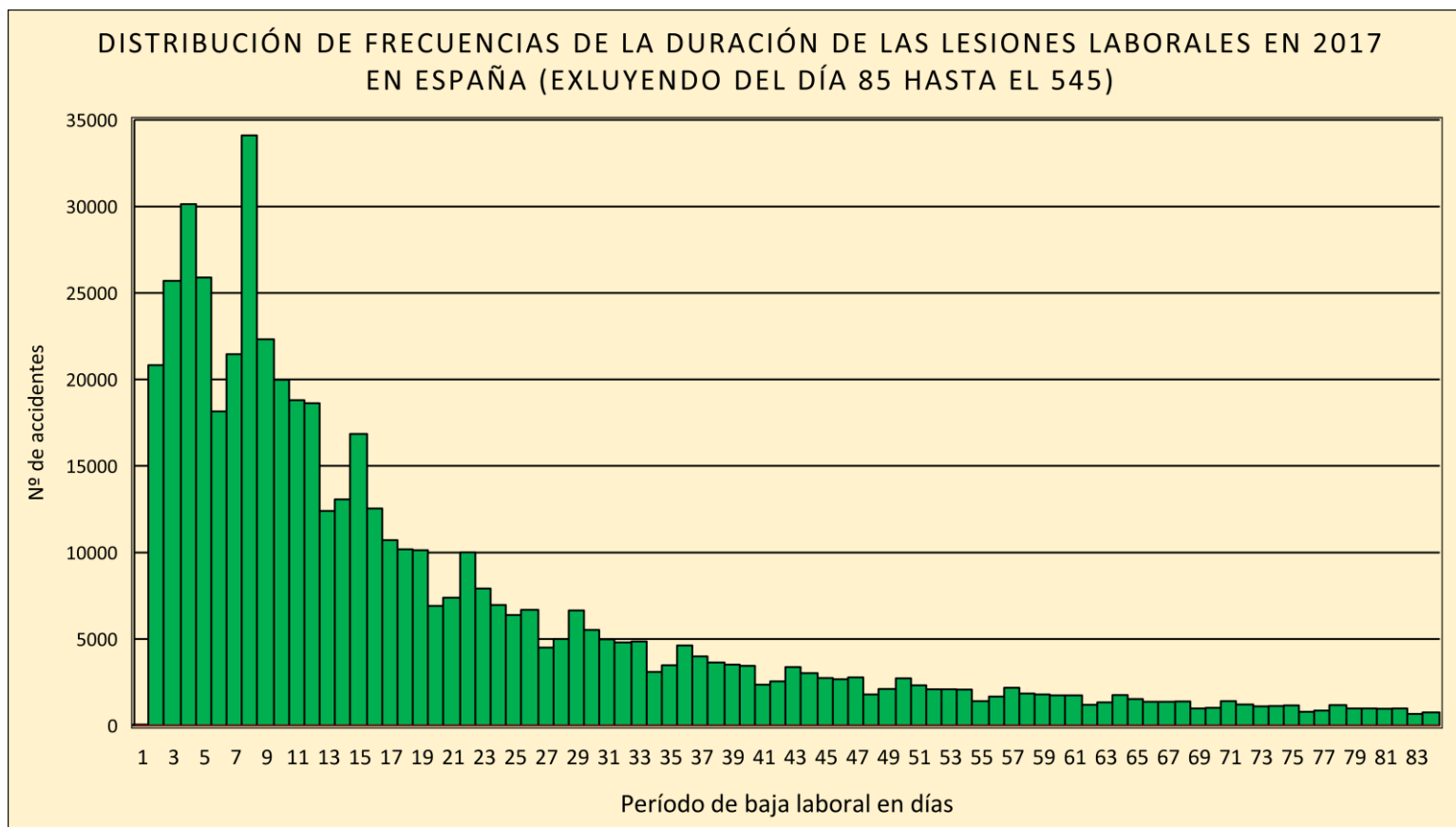


Gráfico 2.1, Elaboración propia. Fuente: Base datos ATR 2017

Tabla 2.1.

Estadísticos descriptivos principales sobre las distribuciones de la duraciones de las bajas laborales para distintos grupos de lesión.

	Media	Mediana	Moda	Desviación estándar	Curtosis	Coef. de asimetría	Nº de datos	Decil 9	Decil 1	Rango D9-D1
Todas	32.2515	15	8	48.5729	18.6132	3.7268	581644	77	4	73
TL01	23.0464	11	4	37.6299	31.3864	4.7090	91896	52	3	49
TL03	21.0454	11	8	33.7154	39.3928	5.2561	51582	46	4	42
TL05	76.7116	54	22	70.8245	4.7953	1.9504	33964	170	15	155
TL08	34.1065	16	8	50.1711	18.6859	3.7125	38772	80	4	76
TL09	30.3067	16	8	44.0905	22.9017	4.0479	157110	68	4	64
TL13	32.9720	16	8	49.2900	19.5343	3.7858	31904	79	4	75
Lumbago	17.4433	10	8	26.1869	73.8859	6.9031	47642	35	4	31

Elaboración propia. Fuente: Base de datos ATR 2017

En primer lugar, el grupo “lesiones superficiales y cuerpos extraños en los ojos”, con casi 92.000 datos representa aproximadamente un 15% de los accidentes registrados. Su media es 23 días, bastante menos que la general, y a su vez tiene una asimetría positiva más marcada aún y más apuntada ($CAF=4.7$ y $Curtosis=31$) con una mediana y moda más a la izquierda también. Por otro lado, su desviación estándar es 37.5, bastante menor y el decil 1 también está en 3 pero el decil 9 en el día 52, lo que muestra una mayor concentración de los datos que en general. En segundo lugar, las bajas laborales clasificadas como “heridas abiertas” son cerca de un 10% del total de las bajas con una duración media de 21, y se presentan más asimétricas hacia la izquierda que antes y muy apuntadas ($CAF=5.25$ y $Curtosis=39.5$). En concreto llama la atención que el primer día de la segunda semana tiene gran cantidad de altas, con hasta casi un 11% de todos sus registros y su día inmediatamente anterior y posterior bastante más bajos. El 80% de los datos se encuentran entre 4º día y el 46º de las distribución, y con una desviación de 33.7, por tanto, en este caso como el anterior están más concentrados los datos, pero más que el otro subgrupo. En tercer lugar, “Fracturas cerradas” se muestra muy distinto a las anteriores, pues suponiendo alrededor del 6% de los datos, tiene una duración media de casi 77 días. La mediana y la moda están también más a la derecha que las anteriores (54 y 22 respectivamente) y está relativamente poco apuntada su distribución y con una asimetría positiva que en comparación es menos notable ($CAF=1.95$ y $Curtosis=4.8$). Por la contra sus datos están menos concentrados (hay más variedad entre los tiempos de recuperación) que se observan en una desviación estándar de 71 bastante alta y un decil 1 en 15 y decil 9 en 170. Su equivalente más grave “Fracturas abiertas” tiene parámetros parecidos, con una media algo mayor, y algo menos apuntada, asimétrica y concentrada. En cuarto lugar, “Dislocaciones y subluxaciones” es casi un 7% de las lesiones y su media está cerca de la total, con cerca de 34 días de período medio de baja. Igualmente, su moda es la misma y su mediana uno más, 16. También presenta similar asimetría, apuntamiento y desviación estándar. La diferencia se observa en que en este grupo concreto los datos están bastante más concentrados pues el 80% de ellos están entre 4 y 80 días de baja laboral. Prosiguiendo en quinto lugar, el grupo “Esguinces y torceduras” es el más representativo de todos siendo alrededor el 27% de las bajas laborales registradas en 2017. Su media es de 30.5 un poco menor que la general, y la mediana y la moda iguales que el grupo anterior (pertenecen al mismo conjunto). Es un poco más apuntado y más asimétrico que todas las lesiones en conjunto ($CAF=4.05$ y $Curtosis=23$). Está más concentrado aún que el anterior pues el decil 1 también está en el 4 y el decil 9 en 68. Y su desviación estándar es de 44. En este conjunto, también se encuentran “otros tipos de dislocaciones, esguinces y distensiones” que, aunque estén el grupo “otros” (similares de otros conjuntos no se han comentado) se comentan brevemente por su representatividad de casi un 15.3% de los registros. Presentan una estadística descriptiva muy parecida a el anterior grupo, pues su media es 30.5 también, y su desviación estándar, curtosis y asimetría son muy parecidos también. El rango inter-decil está entre 4 y 70, un poco menos concentrado. Terminando, el último tipo de lesión considerado como representativo o de interés es “Lesiones internas”, que alberga casi 32.000 datos, casi un 5,5%. El período medio de baja es muy parecido al total, siendo de 33 en este caso. A su vez la mediana es 1 más (16) y la moda igual que la distribución general, y su apuntamiento y asimetría también muy similares. De la misma manera, la desviación estándar es ligeramente mayor por poco y el 80% de los datos están entre 79 días de baja y 4 días, lo que le presenta por lo general con unos valores medios en comparación con otros grupos analizados. A partir de este grupo, y en algunos anteriores no comentados, la cantidad de datos que existen en ellos les hace poco representativos, y revisándoles por encima no se observa nada fuera de lo normal. Sólo con los grupos comentados ya se

Resultados y conclusiones

habría hablado del 85% de los datos, lo que muestra que eran los tipos de lesiones más frecuentes, y en los que centrar la atención a la hora del análisis con el objetivo de optimizar el trabajo.

Mencionado arriba, se ha sacado la distribución de la variable lumbago, que figura dentro de los códigos CIE9, y se encuentra dentro de distintos grupos de “tipo de lesión”, por lo que parece interesante ver cómo se comporta este conjunto saliéndose de los 37 tipos de lesión. Además, figuran hasta casi 48.000 casos de lumbago, por lo que es una enfermedad laboral frecuente. Lo que se obtiene es muy relevante, pues tiene una media de 17.5 días de periodo de baja, con una mediana de 10 y una moda de 8. Su desviación es muy baja, con un valor de 26, pero por la contra presenta una asimetría positiva muy fuerte, de casi 7, valor no observado en los grupos anteriormente comentados, y un apuntamiento medido por curtosis de 74 también muy elevado. A la vez, los datos se encuentran bastante concentrados estando el 80% de las altas entre el día 4 y el día 35. Reflexionando acerca de las hipótesis que se plantean en otros trabajos acerca de esta enfermedad o lesión en las cuales se clasifica como de difícil diagnóstico no parece tener una gran variabilidad de los días de baja. Es más, en comparación con otras lesiones los datos están muy concentrados. Los resultados muestran que la mitad de las bajas laborales por esta causa finalizan el décimo día o antes, y que el 90% de ellas, duran un máximo de 5 semanas, lo que confrontado con grupos anteriores se trata de valores mucho más bajos, más incluso que el grupo de “lesiones superficiales y cuerpos extraños en los ojos” o “heridas abiertas”. Cabe decir eso sí, que aquí se trata un grupo más concreto de lesiones, mientras que los 37 grupos de lesión estipulados, aunque estén bastante afinados no dejan de agrupar a diferentes enfermedades del CIE9, mientras que esta se presenta sin agrupar con otras.

Por tanto, ya se conoce con más detalle cuáles son los tipos de lesión de mayor interés y se ha profundizado en sus distribuciones y parámetros descriptivos. Así pues, una vez realizado este análisis inicial se procede a presentar los resultados de la regresión realizada.

2.2. Presentación de los Resultados

2.2.1. Doble Regresión. Primera parte

Primero, se presenta la regresión con los factores fisiológicos asociados a la lesión, por un lado (Tabla 2.1). En ella se incluyen 8 conjuntos, uno con la variable de género, otro con 7 variables de edad, luego cuatro únicas, que son asistencia ambulatoria, recaída, hospitalización y grado grave, prosiguiendo un conjunto con 8 variables de parte del cuerpo y por último un conjunto de 36 variables de tipos de lesión. Como se observa, se ha obtenido un R^2 del 0.176354 que, dado el estudio previo, las comprobaciones previas y la adecuación, se trata de un nivel de significancia considerable, debido a la dificultad de ajustar este fenómeno ante la cantidad de variables que tiene y lo difícil que son estas de controlar todas. Por tanto, esta primera ecuación tendría un poder explicativo del 17,64 % sobre la variabilidad de este fenómeno, la duración de las bajas laborales. Como ya se ha dicho, es una regresión logarítmica por lo que tendríamos que transformar todas las variables para su correcta interpretación. Para ello además de la tabla 2.2 que figura en la siguiente página se ha sacado la tabla 2.3 (en los anexos) donde figuran todos los coeficientes transformados a porcentajes.

Tabla 2.2 Regresión MCO Variables asociadas directamente a la baja laboral

Variable	Coefficient	t-Statistic	Variable	Coefficient	t-Statistic
C	1.807723	188.3263	TL09	0.226713	50.84956
HOMBRE	-0.032052	-11.02005	TL10	0.231704	46.64707
EDAD2	0.092457	15.77894	TL11	1.073028	35.61877
EDAD3	0.187359	32.85308	TL12	0.476253	15.97823
EDAD4	0.265253	45.72976	TL13	0.253534	37.69963
EDAD5	0.330679	56.84006	TL14	0.252707	15.62697
EDAD6	0.424537	70.13813	TL15	-0.19528	-13.44155
EDAD7	0.530909	56.44627	TL16	0.035465	1.148643
EDAD8	0.520769	8.870711	TL17	0.076152	0.349017
AMBULATORIA	-0.193409	-37.05894	TL18	-0.078355	-2.953647
RECAIDA	0.453569	67.88488	TL19	-0.589752	-6.738041
HOSPITALIZ	0.653989	60.94435	TL20	-0.070712	-1.2471
GRADOGRAVE	1.017092	54.35591	TL21	-0.438655	-8.357765
PARTE2	0.790783	20.71479	TL22	-0.489392	-5.423783
PARTE3	0.792119	96.1992	TL23	0.260337	1.079373
PARTE4	0.458808	62.98091	TL24	-0.536076	-5.634702
PARTE5	0.580542	64.23863	TL25	1.747996	6.158354
PARTE6	0.764029	113.6138	TL26	0.183751	1.887455
PARTE7	0.766139	110.8696	TL27	0.101436	1.955351
PARTE8	0.881026	97.31268	TL28	-0.547516	-4.92788
PARTE9	0.71985	24.74486	TL29	-0.30242	-2.324881
TL02	0.189825	11.08313	TL30	-0.236384	-0.565896
TL03	-0.040373	-6.884685	TL31	-0.406764	-3.68449
TL04	0.037293	5.978621	TL32	0.727349	18.1577
TL05	1.211518	181.3882	TL33	0.007463	0.27813
TL06	1.142	64.41232	TL34	0.363118	3.634283
TL07	1.110553	75.47049	TL35	0.362652	31.34065
TL08	0.318914	50.52231	TL36	1.28478	32.32752
			TL37	0.166573	13.119
R-squared		0.176354	Mean dependent var		2.799242
Adjusted R-squared		0.176274	S.D. dependent var		1.127307
S.E. of regression		1.023136	Akaike info criterion		2.88372
Sum squared resid		608809.8	Schwarz criterion		2.884825
F-statistic		2223.671	Durbin-Watson stat		1.945442
Prob(F-statistic)		0			

Elaboración propia a partir de E-Views 10. Fuente: Base de datos ATR 2017.

A primera vista se ve una regresión donde todas las variables tienen un estadístico-t que muestra que son significativas. Salvo algunas variables de tipos de lesión, la mayoría tienen un estadístico-t alto se podría decir, y las que no, son variables que no reflejan un comportamiento distinto a la referencia utilizada. Desde arriba, se ve primero que la variable del género muestra que el hecho de ser hombre supone una menor duración de la baja, pero tampoco existe una gran diferencia. Luego, se observa que la edad es muy significativa, y que como era de esperar es una variable que según se asciende en tramos de edad los porcentajes aumentan, en concreto un individuo que tenga entre 61 hasta 67 se esperará una baja que dure aproximadamente hasta un 70% más que uno de entre 16 y 24, lo que parece lógico. No obstante, se observa un caso curioso pues del séptimo tramo (61 hasta 67 años) al último tramo (68 hasta 85 años) el porcentaje disminuye ligeramente, descuadrando toda la lógica. Se ha revisado tras ver esto la distribución del último tramo

Resultados y conclusiones

y es cierto que se tienen muchos menos datos que en otros tramos, pero aun así parecen suficientes. Una explicación quizá podría ser que en este tramo se den menos bajas largas debido a que al estar ya en edad de jubilarse, una vez se produzca una lesión larga, el trabajador se jubile.

Las 4 siguientes variables se presentan altamente significativas. Por un lado, está la dummy que representa si el individuo ha tenido asistencia ambulatoria u hospitalaria, y vemos que como parecía razonable esperar, una asistencia ambulatoria supone una menor duración de la baja, concretamente de un 17,5% menos. Es razonable ya que si el individuo sufre un accidente que pueda parecer más grave tenderá a querer recibir directamente una asistencia hospitalaria donde por norma existen más recursos y personal especializado para diagnosticar evaluar y tratar una lesión.

En segundo lugar, que el accidente sea una recaída o no es muy significativo y este hecho implica casi un 57,5% del aumento del tiempo de baja para la estimación realizada. Que el individuo haya sido hospitalizado o no también implica un aumento de la duración de la baja muy grande, concretamente en esta estimación se trata de en torno a un 92%, comprensible también dado que si el individuo no es hospitalizado será porque la gravedad de la lesión no sea tanta. Por último, que la lesión sea de grado grave (recuérdese que engloba a lesión grave y muy grave) a que no lo sea, conlleva un diferencial de un 176,5%, porcentaje muy alto que va acorde con lo pensado.

Después se encuentran las 8 variables de partes del cuerpo, con también considerables niveles de significancia. Así por encima se ve que todas las partes del cuerpo tienen un período mayor de baja que la dejada como referencia (“cabeza”). En concreto las del cuello suponen hasta un 120% más, las de extremidades inferiores o superiores ambas en torno a un 115%, y “espalda y columna” es la que menor diferencial tiene con un 58%. Por último, se presentan los tipos de lesiones, donde se hallan niveles de significancia muy dispares, con estadísticos-t en algunos grupos muy altos frente a otros muy bajos. Simplificando y poniendo atención en algunas concretas que interesen más se observa por ejemplo que el grupo de fracturas, donde se encuentran “fracturas cerradas”, “fracturas abiertas” y “otras fracturas” con un estadístico-t alto muestran un diferencial de un 235%, 213% y 203% más de duración en comparación con el tipo de lesión incluida en la constante (Recuérdese que se trata de “Lesiones superficiales y cuerpos extraños en los ojos”). Son porcentajes altos, aunque pueden parecer razonables, pero el hecho que llama la atención es que las fracturas cerradas salgan de mayor duración que las abiertas para la estimación llevada a cabo. Después se encuentran los grupos 8, 9 y 10, “dislocaciones y subluxaciones”, “esguinces y torceduras” y “otros tipos” que presentan unas duraciones mayores en un 37% para el primero y en torno al 26% el segundo y tercero. Estos tres grupos representan alrededor del 40% de todas las lesiones por lo que son bastante relevantes. Continuando se observa que el siguiente grupo, perteneciente a “amputaciones traumáticas o pérdidas de partes del cuerpo” supone un porcentaje muy alto en comparación de casi un 200% que concuerda con la lógica. Los 3 siguientes, englobando a conmociones y lesiones internas tienen también mayores duraciones como se ve con porcentajes de 60 o 30% aproximadamente al contrario que TL15 (“quemaduras y escaldaduras”) que supone un período de baja un 17.7% menor. Luego a partir de este punto, desde TL16 hasta TL34 los estadísticos “t” disminuyen y algunos casos dejan de ser significativos (excepto TL32, “daños psicológicos debidos a agresiones o amenazas” que si muestra un estadístico t de 18 y con un coeficiente que muestra que este grupo supone un 107% más de tiempo de lesión). Terminando, el grupo TL35 lesiones múltiples es significativo y muestra una duración de casi un 44% más, luego el siguiente grupo, “infartos, derrames cerebrales y otras” tiene el coeficiente más alto de todos implicando

un 261 % más de duración de la lesión en confrontación con “Lesiones superficiales y cuerpos extraños en los ojos”. El último grupo se trata de lesiones no especificadas por lo que no se pone atención en él.

2.2.2. Doble Regresión. Segunda parte

2.2.2.1. Comparación con la regresión anterior

En segundo lugar se presenta la regresión tanto de los factores fisiológicos asociados a la baja como los de riesgo moral. (Se encuentra en la Tabla 2.2 de los anexos del capítulo 2). Para la estimación se añaden 10 conjuntos, uno acerca del tipo de contrato con dos variables, otro con la cuantía diaria de subsidio, otro con la situación profesional del individuo que incluye 3 variables. El siguiente conjunto son 4 variables acerca de los sectores económicos, luego una variable sobre la no predominancia de actividad física en la ocupación, luego la variable del lumbago. Respecto del tiempo se añade un conjunto de los días de la semana con 6 variables, y una de las estaciones del año con 3 variables. Por último, se incluyen también un conjunto con una variable de si el individuo es español y otro conjunto de la comunidad autónoma donde se gestiona el parte y se recibe el tratamiento con 17 variables. Como primera comparación antes de centrar la atención en cada variable al detalle, se habla de las dos regresiones. Vemos que el R cuadrado ha aumentado hasta un 0.1895, lo que supondría que el añadir las variables de riesgo moral al análisis hace que la capacidad explicativa del modelo aumente de un 17.635% hasta un 18.959%, lo que es igual, un aumento de la capacidad explicativa del 1.324%. Es un porcentaje que parece bajo, pero ya se menciona previamente la dificultad de la estimación y el ajuste, por lo que, aunque no se pueda negar que el aumento solo supone un 1,324%, aun así, si muestra una influencia de las variables de riesgo moral, que ayudan a explicar mejor el fenómeno, el cual es muy difícil de estimar en sí. De la misma manera se ve reflejado unos similares resultados de ciertos estadísticos o que mejoran ligeramente. (Mean Dependent Var o Durbin-Watson stat...)

2.2.2.2. Variables “fisiológicas”, comparación resultados

Una vez analizado el resultado de la regresión en sí, se procede a entrar en detalle con cada una de las variables viendo que porcentajes de aumento o disminución suponen respecto de lo que está incluido en la constante que como se ve, ha cambiado un poco respecto a la primera regresión, pues la constante pasa de tener un coeficiente de 1.8077 a 1.7763 que despejando el logaritmo muestra una duración media de la baja para las variables incluidas en la constante en 0.2 días menos aproximadamente. A su vez vemos que el estadístico “t” de la constante ha variado bastante disminuyendo de 188 a 137. Primero, en cuanto a las variables asociadas directamente con el período de baja laboral se advierte que los estadísticos “t” en media aumentan ligeramente, pero centrándonos en grupos o individualmente existen ciertas diferencias. Las variables de edad disminuyen su significancia, sobre todo en los grupos del 4 al 7, pero aun así siguen siendo variables muy significativas en la regresión ya que relativamente no es tanta la caída del estadístico. Las 4 variables de detalles de la lesión (ambulatoria, hospitalización...) siguen mostrándose altamente significativas también, con cambios del estadístico “t” escasos. Las partes del cuerpo por su parte aumentan el valor “t” y destaca sobre todas ellas la variable “Parte 4” referente a “Espalda y columna” con un aumento de su valor “t” de 63 a 76.5, lo que significa que las variables de riesgo moral al ser añadidas aumentan en una gran medida la significatividad de la variable “Parte 4” como explicativa. Por último, los valores “t” de los grupos que engloban a los tipos de lesión presentan resultados dispares. Tratando los mayores cambios o más relevantes, heridas abiertas se muestra más

Resultados y conclusiones

significativa con un 1,3 más en su valor “t”, TL09 y TL10 bajan su significancia al añadir las variables de riesgo moral en 2 puntos absolutos aproximadamente, pero se siguen mostrando significativas a pesar de ello. Lesiones internas también cambia bastante, pues pasa de casi 38 a casi 41 en su valor “t” y por lo demás en general los cambios no suponen una pérdida de significancia ni tampoco una ganancia. Hablando de sus coeficientes transformados a porcentajes, se comprueba que la variable género aumenta un poco el porcentaje y que las variables de edad en general exceptuando la última bajan un poco el porcentaje que suponen a la duración de la baja respecto a los individuos de 16 a 24 años de edad. Destacable que, al añadir las variables de riesgo moral, el fenómeno extraño observado en los dos últimos grupos de edad donde previamente el grupo de 61 a 67 años suponía una baja más larga que el grupo 68 hasta 85 en contra con la lógica, ahora cambie y ya es el último grupo el que muestra que la variación es mayor, aunque por poco. Del conjunto de las 4 dummies siguientes solo la variable grado grave cambia algo más bajando casi un 6% el período de baja mayor que supone que la lesión sea de grado grave o muy grave a que sea leve (de 176% a 170%). Las partes del cuerpo todas aumentan el porcentaje de duración mayor en comparación con la parte del cuerpo “cabeza” donde entre todos los aumentos destaca un aumento del 58.22% al 82.73% de la parte “espalda y columna” acompañando al ya mencionado aumento de su significatividad. Para finalizar, respecto a las variables del tipo de lesión, se observa que las 3 variables pertenecientes al conjunto fracturas de huesos disminuyen el porcentaje, aunque no tanto en términos relativos. El siguiente conjunto, dislocaciones esguinces y distensiones, disminuye poco en términos absolutos, pero en términos relativos algo más, pues por ejemplo el grupo de esguinces y torceduras disminuye de 25.45 a 24.40 por ciento, lo que supone una caída del 4% en términos relativos al añadir las variables que hipotetizan un posible comportamiento diferente entre individuos. TL13, lesiones internas, aumenta relativamente casi un 10% al pasar de 28.86 a 31.68 por ciento. Las siguientes variables “TL” no son mencionadas debido a que no cambian casi ni absoluta ni relativamente o porque presentan muy pocos datos o son poco significativas y por consiguiente no se centra el estudio en ellas.

2.2.2.3. Variables de riesgo moral. Resultados obtenidos

Por otro lado, se tratan las variables denominadas como de riesgo moral. En ellas se muestran valores del estadístico “t” muy dispares, dónde hay valores muy cercanos a cero, y otros alejados que por tanto implican ser significativos. En primer lugar, en aspectos principalmente económicos encontramos bajos valores de “t”. La variable que incluye el tener un contrato temporal u otros tipos de contratos frente a los contratos indefinidos se muestra poco significativa, y de la mano con lo pensado los contratos temporales muestran una menor duración, pero es casi ínfimo el porcentaje pues no llega ni al 1% en negativo, por lo que se podría pensar que no influye apenas esto. En cuanto a la variable continua acerca de la cuantía diaria recibida de subsidio por contingencia profesional se observa una baja significancia y un coeficiente muy bajo, pero hay que recordar que al ser variable continua muestra el porcentaje que dura más una baja por cada euro más percibido de subsidio diario, por lo que dado el porcentaje de 0.01% implica que 1 euro por encima de la media de cuantía diaria de subsidio supone un aumento del 0.01% del tiempo de baja. Es decir, para que el aumento sea del 1% la diferencia deberían ser 100 euros más diarios de subsidio, lo que a priori parece complicado, y aunque existirán casos excepcionales, parece que esta variable apenas influye tampoco. La situación profesional muestra que pertenecer al sector público (sitprof2) frente al privado (variable no incluida), supone un 0.75% menos de duración de la baja. Ello choca en cierta medida con lo pensado quizá, pues se podría asumir que trabajar en el sector público suponga una mayor

seguridad en el empleo lo que llevase a arriesgar más alargando o declarando una baja, pero para la estimación realizada nos dice lo contrario o más bien que no existe diferencia. A pesar de todo, se conforma como variable poco significativa. Por la contra, *sitprof3* y *sitprof4*, que se refiere a autónomos con y sin asalariados muestra un poco más de significatividad y de manera relevante que el que un individuo sea un autónomo con asalariados a su cargo supone una duración de la baja un 32% mayor, y sin asalariados a su cargo un 22.5% mayor. Sería interesante profundizar en ello pues a priori los autónomos podrían considerarse que dada la inestabilidad del autoempleo podrían permitirse menos el alargar o declarar una baja, en mayor medida si no se tiene asalariados a cargo, pero aquí influye un factor no controlable, que es la posibilidad de que dicho autónomo aproveche a darse de baja y a la vez (o a los pocos días, tras ponerse bien) trabaje pero sin darse de alta, agotando el tiempo concedido de baja y así de tal manera sobre el subsidio y siga trabajando de manera ilegal.

En segundo lugar, se encontraban las variables más relacionadas con el tipo de trabajo. El primer grupo hacía referencia al sector económico al que pertenece el individuo, dónde vemos valores “t” altos, y porcentajes relevantes. En confronto con individuos pertenecientes al sector agrario o ganadero los individuos que pertenecen a industrias extractivas suponen una baja en un 14% más de duración y los pertenecientes al sector de la industria y a la construcción suponen un 11,98% y un 7,5% menos respectivamente. Por último, los servicios un 11% menos casi, con lo que se constata que salvo para las industrias extractivas en los demás grupos existe una diferencia con los individuos del sector agrario, que suponen menores períodos de baja que este. Luego fijándose en la variable que evalúa la influencia de si la ocupación en términos generales es predominantemente física o no, se ve que con un valor “t” de 4 tiene un coeficiente en porcentaje de 1.85% más de duración. Se hipotetizó con que en ocupaciones más físicas el período de recuperación pueda ser más largo por necesidad de una total recuperación, pero no parece que sea algo muy determinante.

En tercer lugar, se presentaba la variable que aludía a ese factor que no se ha querido incluir de difícil diagnóstico, pero que siguiendo los precedentes de otros estudios se ha probado a añadir el lumbago, lesión concreta que es tratada en muchos trabajos como una de las principales lesiones de difícil diagnóstico. El resultado que nos va a dar, hay que aclarar que al meter esta variable solo se obtiene el diferencial de duración entre un individuo con lumbalgia a otro individuo con otra cualquier enfermedad. Por tanto, un mayor o menor tiempo de baja no serviría para extrapolar que el lumbago se alarga menos o más que las lesiones de fácil diagnóstico. Aun así, servirá para ver en concreto cómo se comporta esta enfermedad tan concreta y tan citada en otros trabajos de la misma índole. De manera relevante, para la estimación realizada, con los datos utilizados y con la metodología aplicada, la variable muestra un alto valor “t”, lo que quiere decir que tiene un alto poder de explicación dentro de la regresión. Segundo, transformado el coeficiente refleja un 24.09 % menos de duración de la baja que aquellos individuos que tienen una lesión que no sea lumbago.

En cuarto lugar, se presentaban las variables de “tiempo”. Hablando primero de los días de la semana, se observa que el martes y miércoles presentan poca significancia pero que jueves, sábado, domingo y sobre todo viernes tienen un valor “t” considerable. Sus coeficientes son de un 3,5% más de duración si el día de la baja cae en jueves, un 11.5% si cae en viernes, casi un 9% si sábado y casi un 8% si se produjo en domingo. Una posible explicación a esto sería quizá la existencia de muchos trabajos dónde se trabaja de lunes a viernes, lo que explicaría que al no poder volver el sábado o domingo el individuo, aunque esté ya recuperado el sábado no figuraría el alta hasta que fuese lunes.

Resultados y conclusiones

Prosiguiendo, respecto a las variables de los intervalos del año (3 primeros meses, 3 segundos...etc.) se pueden ver un alto valor “t” para el grupo de meses de abril a junio y para el de julio a septiembre. Una hipótesis sería que exista la posibilidad de haya una influencia de la moral de cada persona a la hora de alargar la baja laboral que se note más en meses dónde a priori en España se espera un mejor tiempo meteorológico, sin lluvia o con buenas temperaturas, que haga que el trabajador tenga más incentivo a alargar la baja que con frío o lluvia. Otro factor que podría influir es que el clima afecte al período de recuperación de ciertas enfermedades o incluso que dado un clima se pueda dar con mayor frecuencia un tipo de lesión o no trabajando (como por ejemplo una insolación). Los porcentajes que se obtienen son un 5% menos de duración para Abril-Junio que en Enero-Marzo, un 4% menos para Julio-Septiembre y una casi igual para Octubre-Diciembre.

Para finalizar esta parte del análisis se encontraban en último lugar las variables relativas al espacio. La variable acerca de si el individuo es español o de otro país muestra un valor “t” de 7.75 y el coeficiente dice que un individuo español supone una duración del 3.5% más que si fuera de otro país, lo que puede quizá explicarse porque un individuo de España sienta más seguridad en cuanto a posibilidad de perder el empleo o que se vea con más facilidad para encontrar otro en ese caso. También podría ser que se deba a características culturales que expliquen que se tienda más a realizar este tipo de engaño o no. Concluyendo se tienen las variables de comunidad autónoma, con las que se observan altos valores estadísticos de “t” en general salvo Navarra, La Rioja, Extremadura, Baleares y Ceuta y Melilla que tienen valores por debajo de 10. Comenzado por Galicia, la cual tiene un valor “t” de más de 50 muestra que un individuo de esta región supone una duración un 44% mayor que uno de la Comunidad de Madrid. Asturias y Cantabria también altamente significativas suponen un 31,5% y un 55,75% más respectivamente. Como se comprueba, se presenta muy relevante la pertenencia a una región u otra. País Vasco y Navarra tienen porcentajes más bajos con un 7.35% y un 4.3%. La Rioja, Aragón y Cataluña con significancias entre 8 y 16 de valor “t”, presentan alrededor de un 15% más de duración las dos primeras, y Cataluña de un 6,5%. Castilla y León se presentan también muy significativa y con un porcentaje de casi un 20% positivo. Extremadura y Castilla la Mancha algo menos significativas, presentan porcentajes del 8 y 11 por ciento. Luego, la Comunidad Valenciana tiene un valor “t” de casi 40 y suponen un 25% más. Andalucía y Murcia con valores “t” considerables también suponen un 7,5% más al tratarse de Andalucía y en el caso de Murcia un 20% más. Las dos islas y las dos ciudades autónomas juntas presentan bajos valores estadísticos de “t” (Canarias no tanto), y destaca Baleares con solo un 1,5% más de duración (pero hay que tener en cuenta su bajo valor “t”) y Canarias supone un 7,5% más e igualmente Ceuta y Melilla un 19% más. Como se puede comprobar son muy relevantes los datos referentes a la Comunidad autónoma en la que se tramita el accidente y se lleva a cabo la hospitalización u observación. En ello podrían incidir diferentes factores. Uno de ellos podría plantearse que existen diferencias entre los recursos (tanto humanos como físicos) según la zona a la hora de tratar las lesiones, o a la hora de controlarlas. También podría tener que ver aquí como ya se ha mencionado anteriormente el clima, pues según la zona el clima cambia. Quizá también podría incluso hablarse de cultura de la región o de la mutua que predomine. Los trámites y control que se lleven a cabo en cada Comunidad Autónoma también podrían verse como influyentes por la facilidad que exista para “engañar” al sistema o no. Como vemos, es Madrid la CCAA que menos duración supone, pero a priori parece difícil decir porqué.

2.3. Análisis complementario y resultados de interés

Tras presentar los resultados obtenidos en las dos regresiones, para finalizar se ha realizado dos tipos de análisis a mayores para comprobar la “fortaleza” de los datos y por tanto de su regresión, y a su vez observar ciertos hechos que no se ven a simple vista con la regresión pero que pueden mostrar ser influyentes en su correcta interpretación.

2.3.1.1. Regresión con clasificación Cie-9

En primer lugar, se ha probado a realizar una regresión en la que se eliminan los 37 tipos de lesión y en su lugar de ellos se colocan las enfermedades según la clasificación del Cie-9. Esta clasificación codifica las enfermedades con 3 números enteros y 2 decimales. Por limitaciones de E-Views no se ha podido colocar cada una de las enfermedades (pues no acepta más de 1000 variables para hacer la regresión), por lo que se han truncado los decimales obteniendo así del 0 al 999 todos aquellos conjuntos que presentaban datos, concretamente 629 grupos de enfermedades donde por ejemplo en el grupo 540 están incluidas todas las enfermedades con un código mayor o igual a 540 y menor que 541. (Se truncan debido a que una enfermedad del grupo 399.95 es más similar a las del código 399 y algo que a las del código 400 y algo). Con ello se pasa de tener 37 grupos de lesión a tener 629 pudiendo así analizar cómo cambian la capacidad explicativa de la propia regresión y de las variables de interés junto con el cambio en sus coeficientes, así como ver que códigos del Cie-9 son más explicativos (viendo que enfermedades incluyen) y que coeficientes tienen.

Los resultados son bastante determinantes, y concuerdan con lo pensado, pues la capacidad explicativa de la regresión aumenta de manera considerable. Concretamente el R cuadrado aumenta hasta un 0.305218, lo que supone un aumento de 0.115628. En otras palabras, el aumento del poder explicativo que conlleva este cambio es de un 11.5% lo que es bastante importante.

Visto esto se pasa a observar cómo se han comportado las variables de la segunda regresión con el cambio. En cuanto a sus valores “t” lo primero que se ve es que la constante ha cambiado mucho, pero hay que tener en cuenta que antes en la constante estaban incluidas las “lesiones superficiales y cuerpos extraños en los ojos” y ahora por limitaciones del software de no poder poner otra enfermedad más concreta en la constante, están incluidas aquellas que tenían un código de 0,99 para abajo (que no coinciden con ningún tipo de enfermedad, por lo que se considerarían como enfermedades no clasificadas o desconocidas). Los cambios en el valor “t” son dispares, pero en general bajan su significancia más de los que suben. La variable hombre aumenta su significancia hasta -20 con un cambio un cambio en su coeficiente de -3,55% → -5.72%. Las variables de edad todas se vuelven menos significativas (entre un 10% y un 20% menos en términos relativos) y sus coeficientes acortan la diferencia con los individuos de 16 a 24 años, y más notoria según se avanza en grupo de edad. Parecido ocurre con las variables “ambulatoria”, “hospitalización”, “recaída” y “gradograde” donde las 3 últimas bajan su significatividad alrededor de un 25% y la primera su valor “t” baja un 10%. También acortan sus diferencias porcentuales igualmente con mayor proporción de cambio en las tres últimas, donde gradograde pasa de suponer durar un 170% más a un 105%. Más notable todavía es la caída en las variables de las partes del cuerpo pues dos de ellas bajan su valor “t” alrededor de 10 en términos absolutos y las restantes que tenían valores muy altos caen 55, 65, 90... hasta máximos ahora de 38 como valor “t” (antes 110). Además, los coeficientes caen entre un 60% y un 85% absolutos minorando las diferencias entre cabeza y las demás. Vemos que las variables asociadas directamente a la baja en general

Resultados y conclusiones

bajan su valor “t” y acortan a la vez la diferencia porcentual la gran mayoría. Esto parece normal pues al usar una clasificación más desagregada de los tipos de lesión, se afina más la comparación entre individuos estadísticamente iguales en todo menos en una de las variables, lo que hace que no haya tantas diferencias.

Pasando a las variables de riesgo moral encontramos por la contra más subidas, pero con también algunas bajadas de su valor “t”. Temporal se vuelve menos significativa pero el hecho relevante es que pasa de una influencia negativa a una positiva, y mientras tanto otros contratos permanecen casi inalterados. La cuantía de subsidio diario pasa a ser menos significativa aún, y su coeficiente a ser 0,00%, reflejando según la regresión una influencia nula (antes 0,01%). Pasando a la “situación profesional” del individuo aparece otro cambio de signo en los trabajadores del sector público, que pasan de un -0,74% de duración menos que trabajadores del sector privado a un 1,34% más. Los autónomos por su parte pierden capacidad explicativa y a su vez caen sus porcentajes de mayor duración con respecto a individuos del sector privado. En los sectores económicos casi no hay cambios de los valores “t” ni de sus coeficientes. Otra variable más que cambia de signo es “ocupación no física”, que con el cambio pasa de suponer un mayor tiempo de baja a suponer uno menor (1,85% → -1,29%) lo que parece extraño y difícil de explicar, pero que concuerda más con lo hipotetizado previamente. Lumbago sufre un cambio muy relevante, pues de un valor “t” de -43 pasa a uno de -11, y con una caída del porcentaje de menor duración de -24% → -9%. Respecto de los días de la semana, pasan a tener todos una diferencia positiva con el lunes debido a que el martes y miércoles cambian de signo (pero muy cercanos a 0%) manteniendo casi invariadas sus significancias y sus porcentajes las demás. Las “estaciones” bajan ligeramente su valor “t” y también su diferencial negativo con los 3 primeros meses del año. El hecho de que el individuo tenga nacionalidad española pierde significatividad y pasa de suponer un 3,5% más de duración a un 2,2%. Por último, las Comunidades Autónomas en su mayoría todas aumentan su poder explicativo, pero siendo unas en mayor proporción que otras. A su vez sus coeficientes aumentan el porcentaje en aproximadamente la mitad y le disminuyen en la otra mitad. Destaca el cambio de 55% a 45% de Cantabria y el de Ceuta y Melilla de 19% a 26,5%.

Para acabar con los comentarios de esta regresión, brevemente, se comentan los pequeños grupos de enfermedades más explicativos, es decir con mayor valor “t”. Estos son: “Cuerpo extraño en el ojo”, código 930, que tiene una significancia de 42, y supone un 70% menos de duración que las no clasificadas. El código 836, que incluye luxaciones en la rodilla (y desgarros del menisco), suponen un 200% más de duración y su valor “t” es de 42.5. Luego hay tres grupos muy significativos casi seguidos, las lesiones del código 825 “Fractura tarso y metatarsiano” tienen un valor “t” de 51 y duran un 246% más, el código 824 “Fractura del tobillo” con un valor “t” de 55 dura un 335% más y el código 823 “Fractura tibia y peroné” su valor “t” es 54 y su período de baja de un 350% mayor. Muy cerca también, el grupo correspondiente al código 815 que incluye “Fractura metacarpiano” tiene una significancia de 40.5 y una duración de un 186% más. “Fractura radio y cubito” correspondiente al código 813 tiene un poder explicativo medido por su valor “t” de 53 y supone una baja que dura un 250% más. Siguiendo, el código 812 “Fractura humero” con significatividad de 51 tiene una duración un 345% más. También, con el código 805 referido a “Fractura vértebra sin lesión medula” tiene un valor “t” de 47,5 y una baja que dura un 320% más. Por último, del grupo 722 “Enfermedad disco intervertebral” tiene una significatividad del 46 y una duración en comparación un 253% mayor. Se ha comprobado que todas ellas tienen una cantidad considerable de datos, donde la que menos datos tiene, es 3000. Con esto se ve, que, salvo los dos primeros y

el último, dentro de las enfermedades más explicativas en la regresión con valores “t” por encima de 40 (hecho así para abreviar), se encuentran 7 pertenecientes a fracturas de huesos. Además, se constata que todas estas tienen altas duraciones asociadas.

2.3.1.2. *Interacciones de las principales variables estudiadas*

Como segunda parte de este epígrafe, y para finalizar el análisis acerca de las bajas laborales de España en 2017, se han realizado regresiones con algunas interacciones entre las variables. Al igual que en todo el análisis, se sacan los resultados, pero con el objetivo de no excederse en detalles, se muestran solo los más interesantes. Es decir, de aquellos obtenidos sólo se focalizará en aquellas interacciones con alto valor “t” y que a su vez supongan un cambio del porcentaje relevante (en total se han hecho las interacciones de alrededor de unas 15 variables con todas sus demás posibles). Abajo en la tabla 2.6 se muestran algunas de las interacciones que resultaban ser más significativas.

Como primer ejemplo, en relación con la variable hombre se destaca, por ejemplo, la interacción acerca de hombre-hospitalizado, la cual incrementa un 18% la duración. Si se suma junto con el coeficiente de varón y de hospitalizado, el resultado es lo que dura más la baja con de un hombre-hospitalizado con respecto de la referencia que es mujer-no hospitalizada. Además, usando el coeficiente de hospitalizado, podemos ver que baja es más larga usando el diferencial entre los 3 juntos, o sólo hospitalizado (que se trata de mujer hospitalizada, pues está ya se encuentra incluida en la constante). También se puede hacer en el otro sentido viendo si la interacción compensa o agranda el hecho de ser hospitalizado o no. Rápidamente clarificando con números, hombre-hospitalizado supone un 105,8% de duración más que la referencia (un 18% más de lo pensado antes de saber la interacción) y mujer-hospitalizada en este caso es un 91,35% más de duración. Es decir, la interacción muestra una mayor duración para el hombre, que no era vista previamente.

Dado el estudio, se considera que las interacciones entre una variable fisiológica y una de riesgo moral pueden resultar más interesantes, y por ello, aunque no se encuentren entre los valores “t” más alejados de 0, son las comentadas a continuación.

La variable hombre tiene una interacción curiosa con respecto de las CCAA, pues todas suponen porcentajes negativos, exceptuando Asturias (un 5% más de duración), y Ceuta y Melilla un (23,5% más). Relevante también, es el hecho de que si es un individuo hombre que trabaja por cuenta propia con asalariados a su cargo, la interacción hace caer un 30% el período de baja. Por lo que en este caso a un hombre le dura un 30% menos la baja que a una mujer en la misma situación. Muy relevante es si el individuo es un hombre que trabaja en industrias extractivas ya que supone un 30% menos de duración (con un valor “t” de -2,17), lo que vuelve negativo el efecto de presentarlas ambas juntas frente a mujer del sector agrícola. De igual manera ocupación no física y hombre tienen un “t” de casi 10 y suma un 9% revirtiendo el efecto frente a una mujer de ocupación física.

Prosiguiendo, la interacción de ambulatoria con trabajador por cuenta propia con asalariados a su cargo muestra un porcentaje de un +33% (con un estadístico t de 1.43). Por su parte, recaída en su interacción con las estaciones del año revierte el efecto para Julio-Septiembre y Octubre-Diciembre pues supone un +7% y +8%, durando por tanto más ahora las bajas en estos dos períodos que con respecto de los 3 primeros meses (la referencia). Recaída presentada con las diferentes CCAA muestra porcentajes negativos como para el caso de Andalucía o Murcia, o positivos como es el caso de Aragón, Canarias, Extremadura, Castilla la Mancha y la Comunidad Valenciana.

Con hospitalización se presentan relevantes y significativas sus interacciones con los sectores económicos pues si el individuo hospitalizado trabaja en el sector de las

Resultados y conclusiones

industrias extractivas supone un -24% su interacción, con lo que siendo hospitalizado pasa a durar menos su baja que si perteneciese al sector agrícola. Por la contra si pertenece a la industria es un +17% y a la construcción un +25% pasando por tanto a durar más por pertenecer a esos sectores con respecto de la referencia. Curioso también que su interacción con el domingo (-13,7%) muestre que en este caso las bajas de un individuo hospitalizado un lunes duran más que uno hospitalizado un domingo, al revés que para el caso general. A su vez, su interacción con lumbago supone un -37%.

Relevante también es si la lesión de grado grave es clasificada como lumbago, pues supone un -65%. En el caso de que el individuo con lesión de grado grave tenga una ocupación donde no predomine la actividad física es un -21%, por lo que para este caso al contrario que para el general duran menos las bajas en este tipo de ocupaciones. Destaca entre las CCAA Galicia pues su interacción con grado grave supone un -29%. Luego para la variable lumbago la interacción más relevante es con trabajadores por cuenta propia con asalariados a su cargo, que supone un +60%, dando peso a la hipótesis acerca de los autónomos que puedan alargar la baja, unido a que esta lesión puede ser más difícil de reconocer su recuperación o no. Asimismo si es un lumbago para un trabajador del sector público es un +7,5%, con lo que duraría más la baja para individuos del sector público que del privado en los casos que son lumbago.

Un autónomo sin asalariados a cargo, por ser español dura un 21% más su baja (añadido al +3,5% que de por sí supone español sola). En ciertas CCAA que el individuo sea español aumenta aún más la duración de la baja como es Galicia, Asturias o Cantabria mientras que en otras disminuye un poco la duración como Canarias o la Comunidad Valenciana.

En el caso general, la baja en el sector servicios duraba un poco más que en la industria, pero si se habla de individuos con la lesión “heridas abiertas” la interacción es un +7% con industria y un -7% con servicios, lo que hace que para este tipo de lesión sean más largas las bajas en el sector de la industria. La siguiente lesión analizada es fracturas cerradas, y si es un autónomo con asalariados a cargo el que tiene este tipo de lesión, conforma un -30% de duración volviendo casi igual la duración de una fractura cerrada de un autónomo con asalariados que de un trabajador del sector privado. A su vez, si un lumbago es clasificado como fracturas cerradas es un -60% de tiempo de baja. También es relevante su interacción con el tipo de lesión “Fracturas abiertas” pues supone un -78% y por la contra con “Otros tipos de dislocaciones, esguinces y distensiones” el porcentaje es positivo (un +7%). En las CCAA destaca Ceuta y Melilla donde las bajas por fracturas cerradas duran un 45% más. Esta luego también destaca con fracturas abiertas pues supone un 200% más de duración y con dislocaciones y subluxaciones que es un 60% más. Luego esta misma dentro de las distintas CCAA muestra en general interacciones con porcentajes positivos, pero que con navarra revierte la situación y para este tipo de lesión las bajas duran menos que en la referencia, Madrid. Para el caso de los “Esguinces y torceduras” destacan Baleares, Galicia y Asturias donde sus interacciones muestran duraciones mayores. Luego si son “lesiones internas” destaca Canarias donde supone un 22% más de baja.

Tabla 2.6. Principales interacciones (Más significativas).

Interacción	T-statistic	Coefficiente	Interacción	T-statistic	Coefficiente
PARTE5*HOMBRE	8.4364	19.56%	GRADOGRAVE*HOSPITALIZ	-12.8817	-39.26%
PARTE6*HOMBRE	11.6092	20.65%	PARTE3*HOSPITALIZ	-11.2487	-48.30%
PARTE7*HOMBRE	17.7831	34.07%	PARTE5*HOSPITALIZ	-9.1781	-35.25%
PARTE8*HOMBRE	13.2719	30.72%	PARTE6*HOSPITALIZ	-9.8446	-32.78%
EXTRACTIVAS*HOMBRE	-2.1692	-30.08%	PARTE7*HOSPITALIZ	-6.8937	-24.44%
OCUPACIONNOFISICA*HOMBRE	9.7664	8.97%	PARTE8*HOSPITALIZ	-9.7190	-36.30%
RECAIDA*AMBULATORIA	14.2164	42.21%	TL03*HOSPITALIZ	13.7005	82.36%
HOSPITALIZ*AMBULATORIA	-13.1242	-26.77%	TL35*HOSPITALIZ	10.7008	78.02%
GRADOGRAVE*AMBULATORIA	9.5619	44.25%	RECAIDA*GRADOGRAVE	-11.6703	-68.17%
LUMBAGO*AMBULATORIA	6.1217	17.93%	PARTE5*GRADOGRAVE	-8.5884	-50.47%
PARTE3*AMBULATORIA	6.8330	20.72%	PARTE6*GRADOGRAVE	-8.3870	-47.43%
PARTE5*AMBULATORIA	6.1402	20.05%	PARTE7*GRADOGRAVE	-6.6968	-39.94%
PARTE6*AMBULATORIA	6.0416	14.61%	PARTE8*GRADOGRAVE	-8.1033	-47.31%
PARTE7*AMBULATORIA	4.9797	12.23%	HOMBRE*LUMBAGO	-14.3816	-15.34%
PARTE8*AMBULATORIA	4.5455	13.52%	TL08*LUMBAGO	-13.0663	-21.21%
TL03*AMBULATORIA	-11.6669	-21.84%	TL09*LUMBAGO	-13.2096	-17.58%
TL35*AMBULATORIA	-7.9864	-22.06%	TL10*LUMBAGO	-10.8297	-14.92%
PARTE3*RECAIDA	-11.4538	-40.29%	TL13*LUMBAGO	-13.3511	-24.06%
PARTE5*RECAIDA	-9.1078	-36.67%	BALEARES*LUMBAGO	8.3533	22.81%
PARTE8*RECAIDA	-8.1975	-34.11%	HOMBRE*TL03	9.8593	13.01%
TL05*RECAIDA	-27.7697	-60.58%	PARTE6*TL03	-20.7626	-31.20%
TL06*RECAIDA	-11.1822	-59.08%	PARTE7*TL03	-12.1495	-23.70%
TL07*RECAIDA	-11.4110	-53.48%	GRADOGRAVE*TL05	-12.2093	-38.40%
TL11*RECAIDA	-9.7014	-63.40%	GRADOGRAVE*TL06	-5.2062	-32.46%
RECAIDA*HOSPITALIZ	-17.8097	-52.34%	LUMBAGO*TL06	-3.0057	-78.44%

Elaboración propia. Fuente: Base de datos ATR

2.4. Conclusiones

Para concluir se analiza en retrospectiva el trabajo llevado a cabo, los resultados obtenidos y si con él se han cumplido los objetivos marcados y sacado información interesante tras realizar todo el estudio.

Realizando este estudio se ha constatado que es un campo de estudio muy extenso, con grandes dificultades de tratamiento de los resultados debido a la gran cantidad de factores pero a la vez muy útil ya que son datos muy fiables dada su procedencia, y dado que son datos poblacionales y no una pequeña muestra. Se observa que, en una mayor o menor medida, son muchos los factores influyentes, y que estos no son solo aquellos relacionados directamente con la baja, sino que otros también influyen, ya que ayudaban a ajustar mejor la estimación. Además, en comparación con otros trabajos se ha separado más ciertas variables, permitiendo ver que en la medida que se desagrega más, la estimación gana representatividad y se ajusta más. Dada la amplitud del tema, se sabe que existe una gran variedad de alternativas para analizar el fenómeno, y que claramente sería muy interesante conocer. Asimismo, se plantea también muy interesante la inclusión en el modelo de análisis otros años, lo que permitiría aumentar el tamaño muestral y a su vez, añadiría una nueva componente de riesgo moral que es la influencia del ciclo económico volviendo más complejo este, pero a la vez seguro que más afinado.

Al inicio se planteaba la relevancia en sí de este tema, y con el análisis se refuerza esta idea. Como se ha podido observar en 2017 hubo más de 580.000 accidentes laborales (excluyendo los mortales y fallos de registro). Los cuales presentan una media de 32,25 días de período de baja, es decir, aproximadamente 18.758.019 días de bajas laborales entre todos los accidentes. Yendo más allá, la media de cuantía de subsidio diario percibida por los trabajadores accidentados era de 40,85€/día, con lo que podríamos decir que supone aproximadamente 766.265.000€ de gasto (El cual es cubierto por la Seguridad Social). Esto en el momento actual se perfila como muy importante, pues últimamente los gastos de la Seguridad Social son foco de atención en muchas ocasiones.

Para poder realizar una mejor interpretación de los datos se realizaba un breve estudio de las distribuciones de los datos en general y confrontándolo con las distribuciones de los principales tipos de lesión. Con ello se ha podido constatar de manera relevante el fenómeno acerca de un sesgo en los datos, que muestran una mayor frecuencia de altas en los primeros días de la medida “semana” de la baja laboral, que a pesar de no haber podido explicarlo con total seguridad ni de corregirlo, es muy útil su conocimiento para posteriormente tenerlo en cuenta para las interpretaciones y para una mejor comprensión. De la misma manera al estudiar en detalle las distribuciones de ciertos grupos de lesión, se ha podido ver diferentes amplitudes y concentración de los datos, así como diferentes asimetría y forma en la que se presentan los datos con lo que se consigue entender mejor estos tipos de lesión.

Como se ha podido comprobar las variables consideradas naturales influyen en una gran medida en la duración de la baja laboral, y en detalle se ha podido ver cuanto cada una. Concretamente se observaba que el género no parecía ser muy determinante mientras que la edad ha mostrado tener una gran influencia. Igualmente, las variables que reflejaban si el individuo ha percibido asistencia ambulatoria, si sufría un accidente considerado como recaída, si había sido hospitalizado y si sufría una lesión de grado grave, se han mostrado

como muy influyentes y con una gran capacidad explicativa para el modelo. De la misma manera las partes del cuerpo también enseñaban alta significatividad y con diferentes niveles entre unas y otras. Las variables de tipo de lesión, en las cuales se ha centrado más el estudio eran más dispares entre unas y otras, pero que de manera general se podría decir que todas aquellas dentro del grupo de las más representativas (en cuanto a volumen de datos) mostraban altos poderes de explicación para el modelo y dependiendo de cual se tratase con unos diferenciales en comparación negativos o positivos.

Después al añadir las variables de “riesgo moral” se ha podido ver que mejoraban la estimación en cierta medida. Dentro de ellas no todas han resultado ser muy significativas, pero ello era parte del análisis ya que todas y cada una de ellas se añadían en hipótesis para de esa manera intentar ver de verdad cuales lo eran (para el método utilizado y los datos disponibles). De manera rápida resumiendo, se ha podido observar que las variables económicas puras no han reflejado ser de las más explicativas, mientras que las variables asociadas al tipo de trabajo se mostraban bastante influyentes, sobre todo las de sector económico al que pertenece el individuo. Luego se comprobaba como la enfermedad “lumbalgia” era muy significativa y con una duración menor en comparación. Por otro lado, en las variables de tiempo, ciertos días de la semana, y ciertas estaciones del año sí que figuran como significativamente distintas de la constante también, pero sin suponer diferencias excesivas en las duraciones. En último lugar, las variables espaciales eran en su mayoría muy influyentes y a su vez mostrando en algunos casos altos diferenciales entre unas comunidades y otras principalmente. Por ello se ve que también ejercen una influencia estas variables que a priori no deberían, y que por tanto serían muy interesantes de estudiar más a fondo, sobre todo aquellas más determinantes.

Para finalizar, con los dos estudios complementarios realizados se ha profundizado en niveles más considerables capacitando de un mejor análisis e interpretación. Primero al cambiar las variables que se usan para agrupar los tipos de lesión se observa que mejora mucho la capacidad explicativa del modelo, y que ciertas variables ven modificados sus valores “t” o sus coeficientes, comentando los principales cambios. En general se veía una pérdida de poder explicativo en las variables naturales acompañada de disminuciones a su vez de los coeficientes, que podría parecer normal, y a su vez se veía un aumento del poder explicativo en las variables de riesgo moral, pero con excepciones y donde incluso se observaban cambios de signo en variables como trabajador del sector público, ocupación no física o porcentajes tanto positivos como negativos intensificados. Por otro lado, se analizaban en detalle ciertas variables como podían ser el género, hospitalización, recaída... con ánimo de ver interacciones con todas las demás variables. Los resultados reflejaban que hay ciertas variables que al presentarse juntas aumentaban su influencia (positiva o negativa) o en ciertos casos también hacia disminuir la influencia acercando a 0% el porcentaje de cambio de la duración en total con los 3 efectos (ambas por sí solas más ambas en conjunto). Esto, aunque no parezca a priori relevante, lo es, pues muestra que hay factores que al presentarse en el individuo a la vez no son como en la regresión se creía o que se comportan de una manera un distinta.

De modo general podríamos decir que se cumplen los objetivos del trabajo pues se ha realizado un análisis amplio, meticuloso y riguroso de las bajas laborales de 2017 observando datos relevadores. No obstante, al tratarse de un objetivo no concreto siempre podría existir la opción de completarle más dada la variedad de posibilidades y alternativas en cuánto a métodos a aplicar, y que según se ha ido realizando este estudio salen a la luz algunas de ellas como interesantes a realizar gracias a él.

3. Bibliografía consultada

1. Albertí, C., Jardí, J., López, R. M., Rocamora, I. T., Delclós, J., & Benavides, F. G. (2012). "Incapacidad temporal por enfermedad común y accidente no laboral en Cataluña", 2007-2010. *Archivos de prevención de riesgos laborales*, 15(4), 172-177
2. de Blas, A. M., Román, Á. M., & Caballero, J. C. R. (2013). El papel de las ETTs en la reducción del riesgo moral asociado al seguro por accidentes de trabajo: El caso de España. *Estudios de economía aplicada*, 31(2), 13-26.
3. Bolduc, D., Fortin, B., Labrecque, F., & Lanoie, P. (2002). "Workers' compensation, moral hazard and the composition of workplace injuries". *Journal of Human Resources*, 623-652
4. Boone, J., & Van Ours, J. C. (2002). "Cyclical fluctuations in workplace accidents"
5. Catalina-Romero, C., Martínez-Muñoz, P., Quevedo-Aguado, L., Ruiz-Moraga, M., Fernández-Labandera, C., & Calvo-Bonacho, E. (2013). Predictores de la duración de la incapacidad temporal por contingencias comunes en los trastornos de ansiedad. *Gaceta Sanitaria*, 27(1), 40-46
6. Dionne, G., & St-Michel, P. (1991). "Workers' compensation and moral hazard". *Review of Economics and Statistics*, 73(2), 236-244.
7. García, A. M., Gadea, R., & López, V. (2007). Impacto de las enfermedades laborales en España. Informe. Madrid: Instituto Sindical de Trabajo, Ambiente y Salud.
8. Herrero, H. C., Román, Á. M., & de Blas, A. M. (2008). La duración de las bajas por accidente laboral en España: ¿ Se justifican las diferencias entre comunidades autónomas?. *Revista de Economía Laboral*, 5(1), 73-98.
9. Krueger, A. B. (1990). Workers' compensation insurance and the duration of workplace injuries (No. w3253). National Bureau of Economic Research.
10. Markussen, S. (2012). The individual cost of sick leave. *Journal of Population Economics*, 25(4), 1287-1306.
11. Martín-Román, A. (2008). Moral hazard and gender differences in the workplace accidents insurance. *Empirical Economics Letters*, 7(7), 707-713.

Anexos

12. Martín Román, Á., & Moral de Blas, A. (2005). Los efectos conjuntos de la temporalidad y la indemnización sobre la duración de las bajas causadas por accidentes laborales.
13. Martín-Román, A., Moral, A., & Corrales-Herrero, C. (2013). Shorter or longer? Sick leave spells and the business cycle for self-employed workers in Spain. *Empirical Econ Lett*, 12(10), 1089-1096.
14. Martín-Román, Á. L., & Moral, A. (2014). Differences between Spanish and foreign workers in the duration of workplace accident leave: a stochastic frontier analysis. In *Disadvantaged Workers* (pp. 277-295). Springer, Cham.
15. Martín-Román, A. L., & Moral, A. (2016). Moral Hazard in Monday claim filing: evidence from Spanish sick leave insurance. *The BE Journal of Economic Analysis & Policy*, 16(1), 437-476.
16. Martín-Román, Á., & Moral, A. (2017). A methodological proposal to evaluate the cost of duration moral hazard in workplace accident insurance. *The European Journal of Health Economics*, 18(9), 1181-1198
17. Osterkamp, R., & Röhn, O. (2005). Being on sick leave: Possible explanations for differences of sick-leave days across countries privatization (No. 19). Ifo working paper
18. Velasco, M. S. (2004). Educación y salarios de los trabajadores en España: Nueva evidencia.
19. Waehrer, G. M., Dong, X. S., Miller, T., Haile, E., & Men, Y. (2007). Costs of occupational injuries in construction in the United States. *Accident Analysis & Prevention*, 39(6), 1258-1266.

4. Anexos

4.1. Anexos capítulo 1

Tabla 1.1. Clasificación de la variable “Tipos de lesión”. Elaboración propia.

Código	Tipo de lesión	Código	Tipo de lesión
TL01	Lesiones superficiales y cuerpos extraños en los ojos	TL20	Infecciones agudas
TL02	Lesión desconocida	TL21	Otros tipos de envenenamientos e infecciones
TL03	Heridas abiertas	TL22	Asfixia
TL04	Otros tipos de heridas y lesiones superficiales	TL23	Ahogamiento y sumersiones no mortales
TL05	Fracturas cerradas	TL24	Otros tipos de ahogamiento y asfixia
TL06	Fracturas abiertas	TL25	Pérdida auditiva aguda
TL07	Otras fracturas	TL26	Efectos de la presión (barotrauma)
TL08	Dislocaciones y subluxaciones	TL27	Otros efectos agudos del ruido, la vibración y la presión
TL09	Esguinces y torceduras	TL28	Calor e insolación
TL10	Otros tipos de dislocaciones, esguinces y distensiones	TL29	Efectos de la radiación no térmica (rayos X, sustancias radioactivas, etc)
TL11	Amputaciones traumáticas, pérdidas de partes del cuerpo	TL30	Efectos de las bajas temperaturas
TL12	Conmoción y lesiones intracraneales	TL31	Otros efectos de las temperaturas extremas, la luz y la radiación
TL13	Lesiones internas	TL32	Daños psicológicos debidos a agresiones o amenazas
TL14	Otros tipos de conmoción y lesiones internas	TL33	Choques traumáticos (eléctricos, provocados por un rayo, etc)
TL15	Quemaduras y escaldaduras (térmicas)	TL34	Otros tipos de choques (desastres naturales, choque anafiláctico, etc)
TL16	Quemaduras químicas (corrosión)	TL35	Lesiones múltiples
TL17	Congelación	TL36	Infartos, derrames cerebrales y otras patologías no traumáticas
TL18	Otros tipos de quemaduras, escaldaduras y congelación	TL37	Otras lesiones especificadas no incluidas en otros apartados
TL19	Envenenamientos agudos		

Tabla 1.2 Clasificación variable “Parte del cuerpo lesionada”. Elaboración propia

Código	Parte del cuerpo lesionada (Grupo)	Códigos que incluye
Parte 1	Cabeza (incluye ojos, orejas, zona facial...)	10,11,12,13,14,15,18,19
Parte 2	Parte del cuerpo afectada sin especificar	00
Parte 3	Cuello	20,21,29
Parte 4	Espalda y columna	30,31,39
Parte 5	Tronco y órganos	40,41,42,43,48,49
Parte 6	Extremidades superiores	50,51,52,53,54,55,58,59
Parte 7	Extremidades inferiores	60,61,62,63,64,65,68,69
Parte 8	Todo el cuerpo y múltiples partes	70,71,78
Parte 9	Otras partes no mencionadas anteriormente	99

Tabla 1.3 Grupos de edad utilizados en el estudio. Elaboración propia

Código asignado	Intervalo de años (Ambos incluidos)	Edad4	40 hasta 45
		Edad5	45 hasta 52
Edad1	16 hasta 24	Edad6	53 hasta 60
Edad2	25 hasta 32	Edad7	61 hasta 67
Edad3	33 hasta 39	Edad8	68 hasta 85

4.2. Anexos capítulo 2

Tabla 2.3 Regresión MCO de todas las variables relacionadas directas y de riesgo moral

Variable	Coefficient	t-Statistic	Variable	Coefficient	t-Statistic
C	1.776256	136.962	TL29	-0.426454	-3.278621
HOMBRE	-0.036128	-11.58147	TL30	-0.285736	-0.689785
EDAD2	0.090253	15.40207	TL31	-0.398477	-3.639762
EDAD3	0.184363	31.88407	TL32	0.700602	17.5601
EDAD4	0.26124	44.19783	TL33	0.004798	0.179947
EDAD5	0.326638	54.78994	TL34	0.327137	3.301549
EDAD6	0.415943	66.46644	TL35	0.356477	31.00162
EDAD7	0.520071	54.63345	TL36	1.267646	31.98098
EDAD8	0.528623	9.07676	TL37	0.177327	14.01168
AMBULATORIA	-0.191829	-36.7967	TEMPORAL	-0.007949	-2.695194
RECAIDA	0.454109	68.25401	OTROSCONTRATOS	0.110528	2.202739
HOSPITALIZ	0.648934	60.73011	CUANTDIA	0.0000525	1.176316
GRADOGRAVE	0.996312	53.56078	SITPROF2	-0.007386	-1.63264
PARTE2	0.818188	21.60213	SITPROF3	0.279323	4.422128
PARTE3	0.79689	97.04473	SITPROF4	0.203032	6.850466
PARTE4	0.602844	76.54446	EXTRACTIVAS	0.132907	5.326234
PARTE5	0.583239	64.93274	INDUSTRIA	-0.127651	-18.665
PARTE6	0.767829	114.718	CONSTRUCCION	-0.077212	-10.71328
PARTE7	0.769766	111.8415	SERVICIOS	-0.1151	-18.21711
PARTE8	0.897393	99.48739	OCUPACIONNOFISICA	0.018358	4.241133
PARTE9	0.73041	25.25359	LUMBAGO	-0.275601	-43.0775
TL02	0.170135	9.971822	MARTES	-0.005574	-1.312073
TL03	-0.047689	-8.169193	MIERCOLES	-0.00381	-0.886656
TL04	0.035121	5.653564	JUEVES	0.035027	8.014502
TL05	1.191504	178.9529	VIERNES	0.108604	24.30138
TL06	1.121609	63.61711	SABADO	0.083931	13.90184
TL07	1.097801	75.07661	DOMINGO	0.076342	10.52802
TL08	0.313899	49.89953	ESTACION2	-0.052783	-14.15717
TL09	0.218336	48.97121	ESTACION3	-0.038071	-10.06509
TL10	0.219639	44.27623	ESTACION4	-0.002506	-0.661114
TL11	1.047839	34.89878	ESP	0.034963	7.753864
TL12	0.479854	16.22625	GALICIA	0.366163	53.21848
TL13	0.275185	40.93444	ASTURIAS	0.274221	26.12534
TL14	0.254746	15.86446	CANTABRIA	0.443075	32.41885
TL15	-0.196858	-13.63871	PAISVASCO	0.070923	10.44565
TL16	0.013423	0.438073	NAVARRA	0.041881	3.70462
TL17	-0.029714	-0.137322	LARIOJA	0.138932	8.565798
TL18	-0.072067	-2.736133	ARAGON	0.143209	16.56074
TL19	-0.675992	-7.643337	CATALUNA	0.062533	13.20344
TL20	-0.099354	-1.750401	CASTILLAYLEON	0.177063	24.36651
TL21	-0.441792	-8.376345	EXTREMADURA	0.07964	7.363191
TL22	-0.468251	-5.212533	CASTILLALAMANCHA	0.105079	13.99565
TL23	0.221408	0.925693	COMVALENCIAN	0.226892	39.91357
TL24	-0.519259	-5.479796	ANDALUCIA	0.072377	14.72169
TL25	1.733663	6.159201	MURCIA	0.18321	22.03084
TL26	0.2068	2.132129	BALEARES	0.014446	1.912965
TL27	0.097803	1.896246	CANARIAS	0.071308	10.06883
TL28	-0.56921	-5.16574	CEUTAYMELILLA	0.173351	5.962599

Tabla 2.3 (continuación).

R-squared	0.18959	Mean dependent var	2.79818
Adjusted R-squared	0.189457	S.D. dependent var	1.126911
S.E. of regression	1.014559	Akaike info criterion	2.866951
Sum squared resid	597311.5	Schwarz criterion	2.868816
F-statistic	1429.002	Durbin-Watson stat	1.962106
Prob(F-statistic)	0		

Tabla 2.4 Transformación de los coeficientes de la tabla 2.2

C	6.09655	GRADOGRAVE	176.51%	TL05	235.86%	TL17	7.91%	TL29	-26.10%
HOMBRE	-3.15%	PARTE2	120.51%	TL06	213.30%	TL18	-7.54%	TL30	-21.05%
EDAD2	9.69%	PARTE3	120.81%	TL07	203.60%	TL19	-44.55%	TL31	-33.42%
EDAD3	20.61%	PARTE4	58.22%	TL08	37.56%	TL20	-6.83%	TL32	106.96%
EDAD4	30.38%	PARTE5	78.70%	TL09	25.45%	TL21	-35.51%	TL33	0.75%
EDAD5	39.19%	PARTE6	114.69%	TL10	26.07%	TL22	-38.70%	TL34	43.78%
EDAD6	52.89%	PARTE7	115.14%	TL11	192.42%	TL23	29.74%	TL35	43.71%
EDAD7	70.05%	PARTE8	141.34%	TL12	61.00%	TL24	-41.50%	TL36	261.39%
EDAD8	68.33%	PARTE9	105.41%	TL13	28.86%	TL25	474.31%	TL37	18.12%
AMBULATORIA	-17.59%	TL02	20.90%	TL14	28.75%	TL26	20.17%		
RECAIDA	57.39%	TL03	-3.96%	TL15	-17.74%	TL27	10.68%		
HOSPITALIZ	92.32%	TL04	3.80%	TL16	3.61%	TL28	-42.16%		

Tabla 2.5 Transformación de los coeficientes de la tabla 2.3

C	5.90770	TL05	229.20%	TL29	-34.72%	VIERNES	11.47%
HOMBRE	-3.55%	TL06	206.98%	TL30	-24.85%	SABADO	8.76%
EDAD2	9.45%	TL07	199.76%	TL31	-32.87%	DOMINGO	7.93%
EDAD3	20.25%	TL08	36.88%	TL32	101.50%	ESTACION2	-5.14%
EDAD4	29.85%	TL09	24.40%	TL33	0.48%	ESTACION3	-3.74%
EDAD5	38.63%	TL10	24.56%	TL34	38.70%	ESTACION4	-0.25%
EDAD6	51.58%	TL11	185.15%	TL35	42.83%	ESP	3.56%
EDAD7	68.21%	TL12	61.58%	TL36	255.25%	GALICIA	44.22%
EDAD8	69.66%	TL13	31.68%	TL37	19.40%	ASTURIAS	31.55%
AMBULATORIA	-17.46%	TL14	29.01%	TEMPORAL	-0.79%	CANTABRIA	55.75%
RECAIDA	57.48%	TL15	-17.87%	OTROSCONTRATOS	11.69%	PAISVASCO	7.35%
HOSPITALIZ	91.35%	TL16	1.35%	CUANTDIA	0.01%	NAVARRA	4.28%
GRADOGRAVE	170.83%	TL17	-2.93%	SITPROF2	-0.74%	LARIOJA	14.90%
PARTE2	126.64%	TL18	-6.95%	SITPROF3	32.22%	ARAGON	15.40%
PARTE3	121.86%	TL19	-49.13%	SITPROF4	22.51%	CATALUNA	6.45%
PARTE4	82.73%	TL20	-9.46%	EXTRACTIVAS	14.21%	CASTILLAYLEON	19.37%
PARTE5	79.18%	TL21	-35.71%	INDUSTRIA	-11.98%	EXTREMADURA	8.29%
PARTE6	115.51%	TL22	-37.39%	CONSTRUCCION	-7.43%	CASTILLALAMANCHA	11.08%
PARTE7	115.93%	TL23	24.78%	SERVICIOS	-10.87%	COMVALENCIAN	25.47%
PARTE8	145.32%	TL24	-40.50%	OCUPACIONNOFISICA	1.85%	ANDALUCIA	7.51%
PARTE9	107.59%	TL25	466.14%	LUMBAGO	-24.09%	MURCIA	20.11%
TL02	18.55%	TL26	22.97%	MARTES	-0.56%	BALEARES	1.46%
TL03	-4.66%	TL27	10.27%	MIERCOLES	-0.38%	CANARIAS	7.39%
TL04	3.57%	TL28	-43.40%	JUEVES	3.56%	CEUTAYMELILLA	18.93%