

EL CONSUMO DE MEDICINA PRIVADA Y SU RELACIÓN
CON EL ENTORNO SOCIOECONÓMICO:
UN ESTUDIO A NIVEL AUTONÓMICO

TRABAJO FIN DE GRADO

CURSO 2014-2015



Autor: Diego Vicente Vela

Tutor: Juan Aparicio Baeza

Fecha de entrega: 12 de julio de 2015

Grado de Administración y Dirección de empresas

Facultad de Ciencias Sociales y Jurídicas de Orihuela

Universidad Miguel Hernández de Elche

ÍNDICE

RESUMEN	3
1. INTRODUCCIÓN	4
1.1 SITUACIÓN DEL SISTEMA SANITARIO ESPAÑOL.....	4
1.2 COMO INFLUYE LA CRISIS ECONOMICA EN LA SANIDAD.....	5
1.3 GASTO SANITARIO EN ESPAÑA.....	7
1.4 SITUACION DE LA SANIDAD PRIVADA ESPAÑOLA	9
2. OBJETIVOS	10
3. METODOLOGÍA	11
3.1 DEFINICIÓN DE VARIABLES.....	11
3.2 MUESTRA DE DATOS.....	12
4. ESTUDIO ECONOMÉTRICO	14
4.1 ESTUDIO ECONOMÉTRICO DEL MODELO COMPLETO.....	14
Fase 1: Análisis Preliminar.....	14
Fase 2: Ajuste del Modelo.....	17
Fase 3: Bondad del Ajuste.....	19
Fase 4: Diagnóstico del Modelo	21
4.2 PROCEDIMIENTO SECUENCIAL PARA LA SELECCIÓN DE VARIABLES.....	24
4.3 ESTUDIO ECONOMÉTRICO DEL MODELO SIMPLIFICADO.....	25
Fase 1: Análisis Preliminar.....	26
Fase 2: Ajuste del Modelo.....	27
Fase 3: Bondad del Ajuste.....	29
Fase 4: Diagnóstico del Modelo.....	30
4.4 TRANSFORMACIÓN DE LA VARIABLE RESPUESTA.....	33
5. CONCLUSIONES	33
6. BIBLIOGRAFÍA	36

RESUMEN

La realización del estudio tiene como objetivo averiguar si existe algún tipo de relación estadística entre el consumo en medicina privada de pago y una serie de variables socioeconómicas. Creando un modelo matemático lineal que explica dicha relación.

El estudio se elabora con los datos obtenidos en el año 2013. Datos de las 17 comunidades autónomas, dejando fuera del mismo a las dos ciudades autónomas de Ceuta y Melilla al no considerarse relevantes por su tamaño. Una vez definidas las regiones objeto de estudio se ha elaborado una base de datos que permite obtener la información necesaria para el desarrollo del trabajo.

En la introducción se expone la situación actual del sector sanitario y su segmentación. El sector está dividido en tres grandes bloques como son paciente de seguridad social, paciente de seguro privado y paciente de medicina privada de pago.

Posteriormente se analiza cómo afectan las variables socioeconómicas al segmento de paciente de medicina privada de pago y si existe relación entre ellas. Con el desarrollo econométrico se pretende encontrar el modelo que mejor explique qué variables influyen en el aumento ó disminución de consumo en medicina privada de pago en cada Comunidad Autónoma. Para ello nos basamos en tres estudios realizados de manera secuencial. Tras realizar un amplio trabajo de campo se recogen 46 variables, como eran demasiadas para una base de datos con tan solo 17 registros se hace un primer descarte formando la base de datos con las 10 variables que tenían una mayor correlación simple. A continuación se eliminan las variables irrelevantes para el modelo. Por último, se transforma la variable respuesta para comprobar si se puede mejorar el modelo anterior.

Observamos que los modelos estudiados son buenos ya que nos indican que existe cierta relación lineal entre la situación económica de cada región y el consumo en medicina privada de pago de sus individuos. Esta relación lineal, para algunas variables como mayores de 65 años tiene un efecto positivo sobre el aumento del consumo en medicina privada de pago. Por el contrario, existen otras variables como IDH que tienen el efecto contrario. Permaneciendo todas las demás constantes, al aumentar el IDH el importe de consumo en medicina privada de pago disminuye.

1. INTRODUCCIÓN

Para poder desarrollar este estudio, tenemos que ser conscientes del estado y situación de la sanidad en España para comprender en detalle las repercusiones. Sobre todo de la sanidad privada que es en la que se centra este estudio.

Partimos exponiendo el panorama del sistema sanitario español para poder entender cómo se encuentra en la actualidad la sanidad pública frente a la sanidad privada. Es importante tener una referencia de la crisis económica que sufre España, y como ésta afecta al sector sanitario. Aparecen más grupos sociales vulnerables, se alteran el patrón de riesgos de la enfermedad y se provocan cambios en la demanda de los servicios. Además, los recortes en el gasto público, afecta al gasto sanitario y consecuentemente a la capacidad de respuesta del sistema sanitario a las necesidades de la población.

También es relevante conocer las repercusiones psicológicas que se producen en una parte importante de la sociedad sometida a un período largo de crisis económica. En estas situaciones el consumo se estanca. Por un lado están los individuos que ven reducida su capacidad de compra, pero también hay casos de individuos que gozan de la misma capacidad de compra que por factores psicológicos reducen su consumo en periodos de crisis.

Posteriormente trataremos de exponer los gastos del sistema sanitario público frente al privado, y dentro del privado se comprara el privado de seguro frente al privado de pago que es en el que se centra nuestro estudio.

Por último, explicaremos como el sector sanitario privado funciona como complemento y alternativa al sistema público, contribuyendo de forma importante a su sostenibilidad.

Con toda la información anterior pretendemos ponernos en la situación del panorama actual de sanidad para poder llegar a argumentar nuestro estudio.

1.1 SITUACION DEL SISTEMA SANITARIO ESPAÑOL.

Las tendencias históricas

Para comprender la situación actual hay que tener en cuenta las tendencias de desarrollo del Sistema Nacional de Salud (SNS).

En los años 80 se produce un gran desarrollo del SNS, que se caracteriza por la aprobación de la Ley General de Sanidad (LGS), con la asunción de un modelo similar al de los servicios nacionales de salud, que va ampliando su cobertura poblacional paulatinamente aunque manteniendo algunos espacios de no cobertura y/o cobertura diferenciado que no se finalizó hasta la Ley General de Salud Pública de 2011.

En los 90 empiezan a producirse, al rebufo de las reformas de Thatcher en el NHS, las primeras propuestas privatizadoras de envergadura (la Ley de Ordenación Sanitaria de Cataluña en 1990 y el Informe Abril en 1991).

En 1995 en el Pacto de Toledo se desvincula la financiación sanitaria de la Seguridad Social, con lo que se potencia la atención sanitaria como derecho ciudadano.

En la segunda parte de los 90 comienzan a desarrollarse los primeros pasos en los que se basaría después la privatización. Es el momento de la amplia teorización de la gestión privada, de la empresarización de los centros públicos y de la conveniencia de la instauración de un mercado sanitario.

En el año 2001 finalizan las transferencias a las CCAA, con la consiguiente descentralización de los servicios sanitarios, momento en el que se acaban de aprobar las leyes de ordenación sanitaria que, en lo general, facilitaban los aspectos de desregulación y privatización antes señalados (Marciano Sánchez Bayle y Manuel Martín García, 2013)

La situación Actual

En este momento estamos viviendo una especial contrarreforma sanitaria que se caracteriza por:

- Potente ofensiva privatizadora, se ha pasado de una privatización “a paso de tortuga” que caracterizaba el periodo anterior a grandes operaciones extensivas cuyo mejor ejemplo es el denominado “plan de sostenibilidad” de la Comunidad de Madrid.
- Cambio del modelo sanitario que pasa de ser un modelo basado en el derecho de la ciudadanía a un modelo que pivota sobre el aseguramiento.
- Copagos generalizados y reducción de la cartera básica (aún por desarrollar en concreto)

Resumiendo, observamos como La sanidad pública en España pasa por momentos críticos. Y cada vez más depende del apoyo de la sanidad privada.

1.2 COMO INFLUYE LA CRISIS ECONOMICA EN LA SANIDAD.

España, al igual que los países de la Unión Europea, entra en un estado de contracción económica precedido de una crisis financiera internacional. La actual crisis económica se está revelando como una de las peores desde la gran depresión de los años 20-30. Los efectos generales de la crisis son:

- Aumento de porcentaje de la población en riesgo de pobreza.
- Altas tasas de desempleo.
- Desahucios y demanda de vivienda social.

- Altas tasas de morosidad.
- Incremento de la tasa de suicidios y problemas de salud mental.

La crisis económica que sufrimos, tiene consecuencias sanitarias:

- Impacto en grupos sociales desfavorecidos.
- Cambios en la demanda de servicios.
- Cambios en el patrón de riesgo de enfermedad.

En España se han adoptado una serie de medidas de recorte que han tenido sus consecuencias sobre el conjunto de agentes del sector sanitario:

- Sobre los profesionales sanitarios, bajadas salariales, incremento de la jornada laboral, impulso a la jubilación a los 65 años y reducción de contratos temporales.
- Sobre los proveedores del sistema, bajada de concertos y retrasos en los pagos. En estos momentos el periodo medio de pago se sitúa por encima de los 500 días.
- Sobre la industria farmacéutica, básicamente por la bajada del precio de los medicamentos y el impacto en la demanda del establecimiento del copago en los pensionistas.
- Sobre las inversiones, disminución del presupuesto de inversiones en todas las comunidades. En este punto la sanidad privada juega un papel relevante ya que está comprando cierta aparatología que la sanidad pública no compra y finalmente esta última acaba derivando los pacientes a los centros privados.
- Sobre los ciudadanos, aumento de las listas de espera, pérdida de la cobertura de los inmigrantes irregulares y otros colectivos e introducción del copago farmacéutico para los pensionistas.

La posición de la Unión Europea, al menos desde la Dirección General de Sanidad y Consumidores, ha sido siempre que la crisis económica no puede transformarse en una crisis de salud pública. Pero las medidas de recorte pueden ser eficaces a corto plazo en cuanto a disminución del gasto, pero negativas a corto, medio y largo plazo para el estado de salud e incluso, a medio y largo plazo, pueden tener un impacto económico negativo.

A semejanza de otros países europeos, en España uno de los primeros y más evidentes impactos de la crisis ha sido el recorte de los presupuestos sanitarios de las comunidades autónomas. Según datos del Ministerio de Sanidad, se habría recortado el gasto público en sanidad en 4.989 millones de euros entre 2010 y 2012. El gasto sanitario total en 2012 se habría situado en torno a los 66. 148 millones de euros (6,3% del PIB), del cual el 92% es gestionado por las comunidades autónomas. En cuanto al presupuesto per cápita de las comunidades autónomas para 2013, se ha producido respecto al año anterior una caída media del 2%. Si comparamos el presupuesto de 2013 con 2010, la disminución es de 152€ per cápita (-11,4%). En 2013, el presupuesto medio per cápita de las comunidades autónomas se sitúa en 1.191 euros, con comunidades muy significativamente por debajo de este valor (Comunidad Valenciana

con 965,4€ o Andalucía con 999,2€) (figura 1). No se trata por supuesto de un hecho aislado. Los datos de la OCDE muestran que el gasto sanitario en países de la UE ha caído por primera vez desde 1.975 como consecuencia de las diversas medidas de recorte, con una caída del 0,6% en 2010 (Equipo de Sanidad y Farmacia de PWC, 2013)

Presupuestos sanitarios per cápita, 2012-2013 (euros)



Figura 1. Presupuestos sanitarios per cápita 2013 ("Diez temas candentes de la Sanidad española para 2013" equipo de sanidad y farmacia de PwC)

1.3 GASTO SANITARIO EN ESPAÑA

En 2013, el gasto sanitario público per capital asciende a 1.236€ y el gasto privado per capital a 495€ (Figura 2). El gasto público per capital se ha calculado a partir de los presupuestos del Estado, aplicando una tasa media de desviación del 7% para estimación del gasto real. El gasto privado se compone de gasto de aseguramiento (estimado en 110€ per cápita) y gasto de pago (385€ per capital).

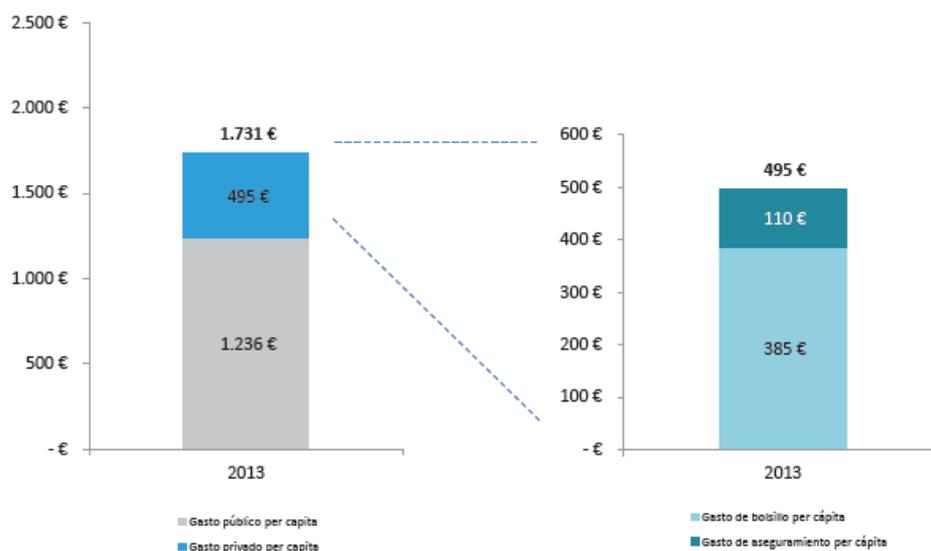


Figura 2. Gasto sanitario público y privado per cápita, 2013 (Informe IDIS, SANIDAD PRIVADA, APORTANDO VALOR ANALISIS DE LA SITUACION 2013, por José Ramón Rubio Presidente del Instituto para el Desarrollo e Integración de la Sanidad)

Respecto al patrón de gasto comunitario, se observa que País Vasco, Navarra y Cantabria son las comunidades que cuentan con un gasto sanitario total más elevado (Figura 3).

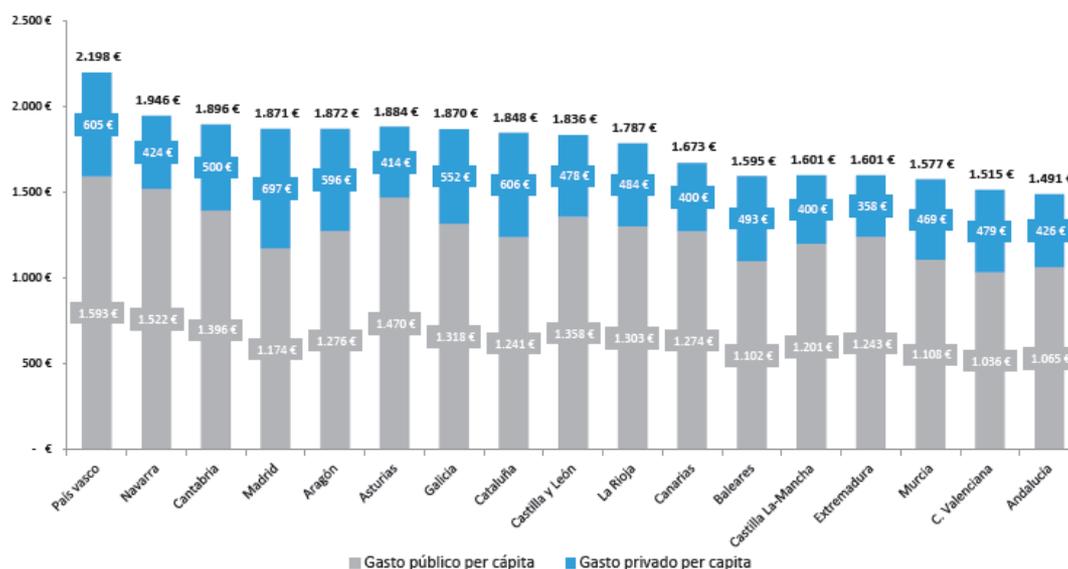


Figura 3. Gasto sanitario público y privado per cápita por CCAA, 2013 (Informe IDIS, SANIDAD PRIVADA, APORTANDO VALOR ANALISIS DE LA SITUACION 2013, por José Ramón Rubio Presidente del Instituto para el Desarrollo e Integración de la Sanidad)

Madrid, Cataluña y País Vasco encabezan el listado de las comunidades con mayor gasto sanitario privado. El gasto privado se ha estimado en base a los gastos de las familias como el seguro privado (fuente ICEA) y con el gasto en salud determinado en la Encuesta de Presupuestos Familiares, realizada por el INE (Figura 4).

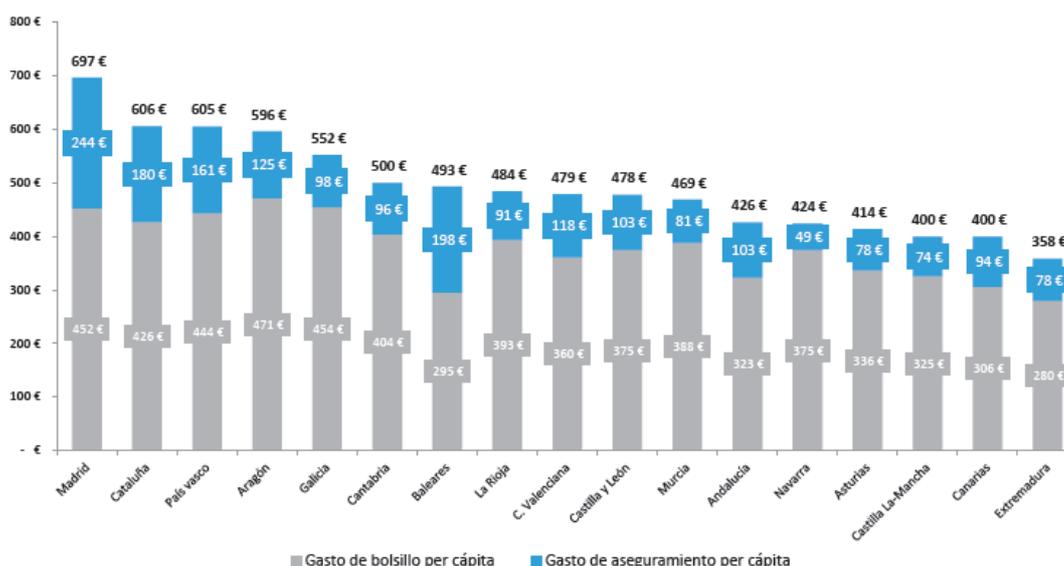


Figura 4. Distribución del Gasto sanitario privado per cápita por CCAA, 2013 (Informe IDIS, SANIDAD PRIVADA, APORTANDO VALOR ANALISIS DE LA SITUACION 2013, por José Ramón Rubio Presidente del Instituto para el Desarrollo e Integración de la Sanidad)

1.4 SITUACION DE LA SANIDAD PRIVADA ESPAÑOLA 2013

La sanidad privada en España, representa un elevado peso en el sector productivo español. Continúa creciendo por encima del PIB, siendo, en dimensión y volumen, comparable e incluso superior a otros sectores clave de nuestro tejido empresarial.

El sector sanitario privado contribuye de forma esencial a descargar de presión al sistema público, ya de por sí desbordado y necesitado de solvencia financiera y sostenibilidad, elementos esenciales de un sistema que gira en torno al paciente.

A pesar de la crisis económica y en términos de crecimiento y desarrollo, el sector sanitario privado continúa registrando un constante crecimiento medio interanual en el volumen de primas.

La sanidad privada mejora el acceso de la población a la asistencia sanitaria, a través de la amplia red de centros privados. Lo que supone una garantía social de primer nivel, que incide en la necesaria complementariedad estratégica del sistema privado de salud respecto a un sistema de titularidad pública altamente reconocido y considerado por todos.

Un aspecto clave, es la colaboración indispensable con el sistema público a través de múltiples formas (concertos, mutualismo o la administración de la gestión de los servicios sanitarios), que supone un equilibrio favorable entre el sistema sanitario público y privado. Estos modelos son muy bien acogidos por parte de los usuarios, especialmente si tenemos en cuenta que el paciente no distingue la titularidad, sino que lo que diferencia es la calidad, la atención personalizada y los resultados en términos de salud; tal es así que, por ejemplo, una gran mayoría de los mutualistas (84,3% el último año) eligen año tras año ser atendidos a través de una entidad aseguradora privada.

El sector sanitario privado es un importante agente inversor en innovación, investigación y en tecnología avanzada. Todo ello, unido a los elevados modelos de calidad acreditados por numerosas certificaciones (los 234.000 profesionales que lo integran, los numerosos ensayos clínicos que se realizan y los 11 hospitales universitarios, junto a las plazas de formación sanitaria), hacen que el sector sanitario privado sea un socio indispensable a la hora de planificar la salud de nuestra sociedad.

En Resumen la sanidad privada, que hoy da cobertura a más de 9 millones de personas en nuestro país, es una realidad sin la cual hoy no sería posible el sistema sanitario español.

El sector sanitario privado es un importante aliado estratégico del sistema público, contribuyendo de manera muy significativa a la sostenibilidad del sistema, la accesibilidad y sobre todo a sustentar e impulsar un sistema de calidad sanitaria. Asimismo, el sector sanitario privado es un agente fundamental en nuestro país en

términos de generación de bienestar, riqueza y contribución al desarrollo económico y social.

2. OBJETIVOS

El objetivo del estudio es determinar si existe algún tipo de relación estadística y elaborar un modelo lineal entre la variable “Consumo de medicina privada de pago en cada CCAA” y variables de carácter socioeconómico como: la población, los hospitales privados, los mayores de 65 años, el PIB per cápita, la tasa de paro, la presión fiscal per capital, la formación superior, el abandono de estudios, la esperanza de vida y el índice de desarrollo humano. Que definiremos en el siguiente apartado.

Posteriormente, se crea un modelo matemático de carácter lineal que permita pronosticar el consumo en medicina privada de pago a partir de una serie de variables socioeconómicas.

3. METODOLOGÍA

El programa utilizado para los cálculos realizados en el estudio es el software libre R, uno de los más utilizados en la investigación por la comunidad estadística.

El manual utilizado para la elaboración de este estudio es Modelos Lineales Aplicados en R de los profesores Juan Aparicio, María Asunción Martínez Mayoral y Javier Morales del Departamento de Estadística, Matemáticas e Informática de la Universidad Miguel Hernández.

Para el análisis preliminar y la determinación de las relaciones lineales entre variables se utilizan los gráficos de dispersión y el cálculo de correlaciones simples y parciales.

Posteriormente, para el ajuste del modelo se realiza la estimación del modelo, el intervalo de confianza y el contraste de hipótesis.

Seguidamente, para la bondad del ajuste se utiliza la covarianza del error estándar residual, la tabla de ANOVA, y el coeficiente de determinación.

Finalmente, para el diagnóstico del modelo además del histograma de residuos, el gráfico qq-plot y otra serie de gráficos se utiliza el test de Shapiro-Wilks para la normalidad, el test de Breusch-Pagan para la homocedasticidad y el test de Durbin-Watson para la incorrelación.

Una vez finalizado este análisis mediante AIC se eliminan las variables menos relevantes para crear el mejor modelo posible y se vuelve a hacer el estudio completo para comprobar si efectivamente el modelo resultante final ha mejorado el previo.

3.1 DEFINICIÓN DE LAS VARIABLES

Pasamos a definir las variables utilizadas en el estudio.

- **Consumo sanitario privado de pago:** Es el importe gastado por los individuos en sanidad privada de pago. El individuo disfruta de la sanidad privada pagando por cada acto médico, o bien a través de la contratación de un seguro privado de salud que proporcione servicios a cambio del pago de una prima periódica. En nuestro estudio quedan excluidas las aseguradoras.
- **Población:** Conjunto de personas que habitan en una región. Nosotros tomaremos las cifras de las personas físicas censadas en cada comunidad autónoma.
- **Hospitales privados:** Entendemos por hospital privado el Hospital que funciona como una organización lucrativa. Muchos hospitales comerciales son propiedad de médicos que dedican el hospital principalmente para sus propios pacientes, aunque también aceptan pacientes de otros médicos. Algunos hospitales comerciales son propiedad de grupos de investigadores o de grandes corporaciones. Para el estudio se utiliza el Número de hospitales privados en cada región.
- **Mayores de 65 años:** Conjunto de individuos cuya edad les permite abandonar la vida laboral y obtener una retribución por el resto de su vida. Nos interesa para nuestro estudio este colectivo en concreto por la relación; a mayor edad, mayor probabilidad de enfermedad y mayor gasto en sanidad.
- **PIB (Producto Interior Bruto) per cápita:** Suma del valor de los bienes y servicios finales producidos en una región durante un año dividido por el número de sus habitantes en ese año. Se utiliza como una medida indirecta de la calidad de vida de la población.
- **Tasa de paro:** Nivel de desocupación entre la población económicamente activa. Se calcula con la Población de 15 años y más que no está trabajando y busca trabajo dividido por la población económicamente activa de 15 años y más (ocupados más desocupados). Con esta fórmula sabemos qué porcentaje de la población económicamente activa está desempleada (es decir, no tiene un empleo y lo está buscando).
- **Presión fiscal per cápita:** Ingresos que la población aporta al Estado en concepto de impuestos. Este indicador se calcula de la suma de los impuestos indirectos, más los impuestos directos, más las tasas y precios públicos, divididos por el número de habitantes.
- **Formación superior:** Porcentaje de la población con estudios universitarios. 1º y 2º ciclo de educación superior y doctorado.

- **Abandono de estudios:** Porcentaje de la población que ha abandonado los estudios en preescolar, primaria y 1ª etapa de educación secundaria.
- **Esperanza de vida:** La esperanza de vida es la media de la cantidad de años que vive una determinada población en un cierto periodo de tiempo. Esta estimación es muy difícil de medir con exactitud ya que dependerá del sexo, de la tasas de mortalidad, de nivel de educación, de las condiciones sanitarias, de las medidas de prevención, y el nivel económico.
- **IDH (Índice de desarrollo humano):** El índice de esperanza de vida, junto al de educación y al de producto bruto interno, componen el índice de desarrollo humano (IDH) que tiene como objetivo evaluar la calidad de vida de los países del mundo. Indicador elaborado por el Programa de las Naciones Unidas para el Desarrollo (PNUD). Se basa en un indicador social estadístico compuesto por tres parámetros: vida larga y saludable, educación y nivel de vida digno.

3.2 MUESTRA DE DATOS

Se elabora una base de datos del año 2013 con las 10 variables socioeconómicas con una mayor correlación simple de las 46 con las que empezamos a trabajar. A continuación enumeramos algunas de las descartadas:

- Trabajo temporal.
- Riesgo de pobreza.
- Renta media.
- Gasto público sanidad pública pc.
- Renta disponible pc.
- Valoración sistema sanitario público.
- Salario medio.
- Tasa Inflación.
- Satisfacción situación económica.
- Dificultad en llegar a fin de mes.
- Gastos vivienda.
- Retraso en los pagos.
- Salarios bajo.
- Satisfacción con el empleo.
- Acceso a la sanidad.
- Confianza en la política.

Las fuentes de información han sido extraídas del Instituto Nacional de Estadística, Ministerio de Sanidad, Asuntos Sociales e Igualdad e Informe de la fundación Idis “Sanidad privada, aportando valor análisis de la situación 2013”.

CCAA	Gto_pri_pago	Poblacion	Hosp_priv	mayores65	PIB_pc	Tasa_paro	Presion_fiscal_pc	Forma_sup	Aban_estu	Espe_vida	IDH
Andalucia	2.710.171.552	8.390.624	49	1.282.926	16.666	36,22	6.018	27,30	28,70	81,61	0,860
Aragon	628.590.948	1.334.588	6	267.851	24.732	21,39	8.862	34,80	18,90	83,16	0,901
Asturias	357.248.976	1.063.241	6	240.717	20.591	24,13	7.436	37,70	19,10	82,23	0,894
Baleares	329.000.225	1.115.255	8	158.923	23.446	22,26	8.467	26,80	29,80	82,69	0,876
Canarias	645.671.934	2.110.039	20	292.873	18.873	33,73	6.710	27,30	27,50	82,33	0,869
Cantabria	237.809.752	588.638	1	111.782	21.550	20,44	8.081	37,20	12,10	83,08	0,900
Castilla y Leon	940.340.250	2.507.574	10	576.992	21.879	21,75	7.962	33,70	19,20	83,57	0,902
Castilla-La Mancha	677.506.375	2.084.635	8	368.146	17.780	29,97	6.569	26,20	27,40	83,17	0,871
Cataluna	3.174.245.706	7.451.281	83	1.268.953	26.666	23,12	9.926	34,80	24,70	83,08	0,895
ComValenciana	1.788.513.480	4.968.093	22	852.524	19.502	28,05	7.042	30,80	21,70	82,50	0,876
Extremadura	307.649.720	1.098.749	6	211.399	15.026	33,87	5.426	24,50	29,20	81,89	0,859
Galicia	1.250.297.840	2.753.960	21	628.178	20.399	22,04	7.310	32,10	20,20	82,69	0,886
Madrid	2.887.708.220	6.388.735	31	983.865	28.915	19,76	10.522	46,00	19,70	84,26	0,919
Murcia	567.654.864	1.463.028	12	208.481	17.901	28,98	6.464	28,20	26,30	82,35	0,868
Navarra	238.744.500	636.652	2	113.897	28.358	17,93	9.925	40,30	12,90	83,63	0,918
Pais Vasco	963.552.372	2.170.163	17	430.560	29.959	16,58	10.236	46,80	9,90	83,22	0,924
Rioja	124.374.282	316.474	2	59.085	25.277	20,04	8.917	35,20	21,70	83,63	0,900

Tabla 1. Base de datos con las variables objeto de estudio.

4. ESTUDIO ECONOMÉTRICO

4.1 ESTUDIO ECONOMÉTRICO DEL MODELO COMPLETO

Con nuestro estudio tratamos de describir la relación entre la variable respuesta (en nuestro caso aleatoria) y las variables explicativas (que sólo son de carácter informativo) a través de un hiperplano, lineal en las variables explicativas, lo más próximo posible a los valores observados de la respuesta. En presente estudio la variable respuesta es "consumo en sanidad privada de pago" mientras que las variables explicativas son "diez indicadores socioeconómicos".

La idea básica de la regresión múltiple es explicar todo lo posible sobre la variable respuesta con una variable explicativa, y a continuación utilizar información adicional que aporta otra variable explicativa y así sucesivamente con todas las variables explicativas para completar la predicción sobre la variable respuesta.

La formulación del modelo para nuestro análisis es:

$$\text{Consumo sanidad privada de pago} = \beta_0 + \beta_1 * \text{Población} + \beta_2 * \text{Hospitales Privados} + \beta_3 * \text{Mayores 65} + \beta_4 * \text{PIB pc} + \beta_5 * \text{Tasa paro} + \beta_6 * \text{Presión fiscal pc} + \beta_7 * \text{Formación superior} + \beta_8 * \text{Abandono estudios} + \beta_9 * \text{Esperanza vida} + \beta_{10} * \text{IDH}$$

Definición de parámetros:

β_0 : término de interceptación o intercepto.

$\beta_1, \beta_2, \dots, \beta_{10}$ son los coeficientes de regresión, son las pendientes de sus respectivas variables, por tanto el valor de β_1 me indica que siendo las demás variables iguales, si aumento en una unidad la variable Población implica que el consumo en medicina privada de pago aumentará en el valor de β_1 . El valor del coeficiente β_1 cuantifica el peso que tiene la variable Población explicando la respuesta media de consumo en medicina privada de pago. Cabría decir lo mismo para el resto de parámetros.

El modelo de regresión múltiple se va a analizar en cuatro fases:

1. Análisis preliminar.
2. Ajuste del modelo.
3. Bondad del ajuste.
4. Diagnóstico del modelo.

FASE 1: ANÁLISIS PRELIMINAR

Para modelizar la base de datos, los trasladamos a una inspección gráfica, con el fin de descubrir qué tipo de relaciones hay entre las variables. Estudiando el gráfico de dispersión obtendremos una primera aproximación de las relaciones entre las variable.

Para conseguir averiguar si la asociación entre las variables se puede catalogar como lineal o no, realizaremos un análisis numérico de las correlaciones simples y parciales. El coeficiente de correlación parcial medirá el grado de relación lineal que existe entre la variable respuesta y cada una de las variables explicativas, teniendo siempre en cuenta la información de las variables restantes a la hora de calcular el coeficiente.

Nos centramos en el análisis, la figura 5 nos muestra todos los gráficos de dispersión posibles. Gráficamente observamos una relación lineal “positiva” importante entre el consumo en sanidad privada de pago y la población, el número de hospitales privados y los mayores de 65 años. Para asegurarnos de que realmente existen relaciones lineales entre la variable respuesta y las variables explicativas calcularemos los coeficientes de correlación.



Figura 5. Relación entre el consumo en sanidad privada de pago y el resto de variables socioeconómicas.

Analizando los resultados de la Tabla 2 comprobamos que se confirman unas correlaciones simples próximas a 1 y por lo tanto altas entre la variable respuesta consumo sanitario privado de pago y las variables explicativa Población y personas mayores de 65 años. La siguiente más importante es el número de hospitales privados. Esto nos indica que tienen una relación positiva, es decir, al aumentar nuestras variables explicativas el número de consumo en medicina privada de pago también aumenta.

VALORES DE LAS CORRELACIONES SIMPLES	
	Gto_pri_pago
Gto_pri_pago	1,0000
Poblacion	0,9774
Hosp_priv	0,8949
mayores65	0,9750
PIB_pc	0,1662
Tasa_paro	0,1270
Presion_fiscal_pc	0,2153
Forma_sup	0,1679
Aban_estu	0,1555
Espe_vida	0,0490
IDH	0,0320

Tabla 2. Valores correlación simple. Modelo completo.

En la tabla 3, obtenemos los valores estimados de las correlaciones parciales. Observamos como la variable explicativa Formación superior es la única que tiene un grado de relación lineal con la variable respuesta consumo en sanidad privada de pago. A pesar de que esta variable en las correlaciones simples es poco significativo, en presencia del resto de variables explicativas parece indicar que sí que es relevante y debemos tenerla en cuenta en nuestro estudio.

VALORES DE LAS CORRELACIONES PARCIALES	
	Gto_pri_pago
Gto_pri_pago	1,0000
Poblacion	-0,0064
Hospitales_priv	0,1631
mayores65	0,6057
PIB_pc	-0,4363
Tasa_paro	0,1785
Presion_fiscal_pc	0,4732
Formacion_superior	0,7660
Abandono_estudios	0,1077
Esperanza_vida	0,5943
IDH	-0,6202

Tabla 3. Valores correlación parcial. Modelo completo.

FASE 2: AJUSTE DEL MODELO

En esta segunda fase ajustaremos el modelo para ver que variables contienen suficiente información para explicar la variable respuesta Consumo sanitario privado de pago, en presencia del resto de covariables.

Veremos los valores que estima el modelo para $\beta_0, \beta_1, \beta_2, \beta_3, \beta_4, \beta_5, \beta_6, \beta_7, \beta_8, \beta_9, \beta_{10}$ con unos intervalos de confianza al 95% para cada uno de los coeficientes y el contraste de hipótesis asociado a la cuestión de si cada coeficiente es igual o no a cero. La variable que rechace la hipótesis de coeficiente cero, será una variable útil.

Los valores de los parámetros, por ejemplo β_1 , representan la estimación del incremento que experimenta la variable respuesta cuando la variable explicativa asociada a dicho coeficiente aumenta su valor en una unidad y las demás variables se mantienen constantes.

En la Tabla 4 mostramos las estimaciones máximo verosímiles de los parámetros, así como el p-valor asociado a los contrastes anteriormente explicados. Además, en la tabla aparecen otros valores que más tarde analizaremos y que se encuentran relacionados con la bondad del ajuste. En definitiva, el modelo ajustado se correspondería con la siguiente expresión:

$$\text{Consumo sanidad privada de pago} = 5.023 \cdot 10^9 - 2.757 \cdot \text{Población} + 2.411 \cdot 10^6 \cdot \text{Hospitales Privados} + 2.020 \cdot 10^3 \cdot \text{Mayores 65} - 9.335 \cdot 10^4 \cdot \text{PIB pc} + 1.271 \cdot 10^7 \cdot \text{Tasa paro} + 4.4452 \cdot 10^5 \cdot \text{Presión fiscal pc} + 7.302 \cdot 10^7 \cdot \text{Formación superior} + 4.439 \cdot 10^6 \cdot \text{Abandono estudios} + 2.271 \cdot 10^8 \cdot \text{Esperanza vida} - 3.164 \cdot 10^{10} \cdot \text{IDH}$$

RESUMEN DEL MODELO											
	β_0	β_1	β_2	β_3	β_4	β_5	β_6	β_7	β_8	β_9	β_{10}
Parámetro	$5,023 \cdot 10^9$	$(-),2,757$	$2,411 \cdot 10^6$	$2,020 \cdot 10^3$	$(-),9,335 \cdot 10^4$	$1,271 \cdot 10^7$	$4,452 \cdot 10^5$	$7,302 \cdot 10^7$	$4,439 \cdot 10^6$	$2,271 \cdot 10^8$	$(-),3,164 \cdot 10^{10}$
p-valor	0,681	0,987	0,699	0,111	0,279	0,672	0,236	0,026	0,799	0,12	0,1
Residual Standard Error	121900000										
CV	11,6231										
Multiple R-squared	0,9943										
Ajusted R-squared	0,9849										
p-valor	$6,303 \cdot 10^{-6}$										

Tabla 4. Resumen del modelo completo

Por otra parte, en la Tabla 5 se muestran los intervalos de confianza al 95% para cada uno de los coeficientes del modelo lineal.

INTERVALO DE CONFIANZA		
	Extremo inferior	Extremo superior
Intercepto	$-2,353383 \cdot 10^{10}$	$3,358029 \cdot 10^{10}$
Población	$-4,290424 \cdot 10^2$	$4,235277 \cdot 10^2$
Hospitales privados	$-1,215736 \cdot 10^7$	$1,697972 \cdot 10^7$
Mayores 65	$-6,304370 \cdot 10^2$	$4,670581 \cdot 10^3$
PIB pc	$-2,856453 \cdot 10^5$	$9,895138 \cdot 10^4$
Tasa paro	$-5,722265 \cdot 10^7$	$8,263423 \cdot 10^7$
Presión fiscal pc	$-3,827391 \cdot 10^5$	$1,273138 \cdot 10^6$
Forma superior	$1,181869 \cdot 10^7$	$1,342310 \cdot 10^8$
Abandono estudios	$-3,649202 \cdot 10^7$	$4,536915 \cdot 10^7$
Esperanza vida	$-7,987878 \cdot 10^7$	$5,341377 \cdot 10^8$
IDH	$-7,161521 \cdot 10^{10}$	$8,336631 \cdot 10^9$

Tabla 5. Intervalos de confianza. Modelo completo.

$$IC_{95\%} \beta_1 = (-4,290424 \cdot 10^2, 4,235277 \cdot 10^2)$$

$$IC_{95\%} \beta_2 = (-1,215736 \cdot 10^7, 1,697972 \cdot 10^7)$$

$$IC_{95\%} \beta_3 = (-6,304370 \cdot 10^2, 4,670581 \cdot 10^3)$$

$$IC_{95\%} \beta_4 = (-2,856453 \cdot 10^5, 9,895138 \cdot 10^4)$$

$$IC_{95\%} \beta_5 = (-5,722265 \cdot 10^7, 8,263423 \cdot 10^7)$$

$$IC_{95\%} \beta_6 = (-3,827391 \cdot 10^5, 1,273138 \cdot 10^6)$$

$$IC_{95\%} \beta_7 = (1,181869 \cdot 10^7, 1,342310 \cdot 10^8)$$

$$IC_{95\%} \beta_8 = (-3,649202 \cdot 10^7, 4,536915 \cdot 10^7)$$

$$IC_{95\%} \beta_9 = (-7,987878 \cdot 10^7, 5,341377 \cdot 10^8)$$

$$IC_{95\%} \beta_{10} = (-7,161521 \cdot 10^{10}, 8,336631 \cdot 10^9)$$

A continuación, estudiaremos si cada coeficiente puede ser o no igual a cero mediante un contraste de hipótesis. En la Tabla 4 aparecen los p-valores asociados a cada par contraste-coeficiente. Por ejemplo, para el caso de beta 1, el contraste de hipótesis a resolver sería el siguiente:

$$H_0: \beta_1 = 0$$

$$H_1: \beta_1 \neq 0$$

En este caso el p-valor es igual a 0,987. Como es mayor que 0,05, no podemos rechazar la hipótesis H_0 y, por tanto, nos indica que en presencia de las demás variables ésta no es relevante y asumiremos que el parámetro es nulo. Siguiendo estos parámetros para el resto de coeficientes obtendríamos:

p-valor $\beta_2 = 0,699 > 0,05$ No rechazamos H_0 , asumiremos que el parámetro es nulo.

p-valor $\beta_3 = 0,111 > 0,05$ No rechazamos H_0 , asumiremos que el parámetro es nulo.

p-valor $\beta_4 = 0,279 > 0,05$ No rechazamos H_0 , asumiremos que el parámetro es nulo.

p-valor $\beta_5 = 0,672 > 0,05$ No rechazamos H_0 , asumiremos que el parámetro es nulo.

p-valor $\beta_6 = 0,236 > 0,05$ No rechazamos H_0 , asumiremos que el parámetro es nulo.

p-valor $\beta_7 = 0,026 < 0,05$ Rechazamos H_0 , el parámetro es distinto de cero.

p-valor $\beta_8 = 0,799 > 0,05$ No rechazamos H_0 , asumiremos que el parámetro es nulo.

p-valor $\beta_9 = 0,120 > 0,05$ No rechazamos H_0 , asumiremos que el parámetro es nulo.

p-valor $\beta_{10} = 0,100 > 0,05$ No rechazamos H_0 , asumiremos que el parámetro es nulo.

Apreciamos que el p-valor relacionado con β_7 es igual a $0,026 < 0,05$, está asociado al porcentaje de personas con formación superior y es la variable más importante de la base de datos coincidiendo con la información obtenida en la Tabla 3, los valores de correlación parcial.

En el resto de betas, como los p-valores son superiores a 0,05 no se puede rechazar H_0 . En consecuencia es posible que tanto beta 1, 2, 3, 4, 5, 6, 8 y 9 sean cero, probablemente la información contenida en estas variables, en presencia de las demás, no explican lo suficientemente bien la variable consumo sanitario privado de pago y es probable que la información contenida en alguna de ellas se solape. Lo que significa que estas variables asociadas no son lo suficientemente útiles. En definitiva, el método del contraste de hipótesis me dice que sólo el coeficiente de Formación superior puede ser distinto de cero (en presencia del resto de variables explicativas).

FASE 3: BONDAD DEL AJUSTE

Al realizar un ajuste en un modelo de regresión lineal debemos verificar si dicho modelo proporciona un buen ajuste a la hora de explicar la variable respuesta, en nuestro caso el Consumo sanitario privado de pago. Para ello contamos con varias medidas para cuantificar la bondad del ajuste de diversos modos, en concreto tres:

1.- Error estándar residual: es una medida de bondad del ajuste relativa a la escala de medida utilizada. Se prefieren modelos con menor error residual estimado. No obstante, una vez estimado el error estándar residual del modelo es habitual determinar si éste es grande o pequeño a través del cálculo del coeficiente de

variación: $cv = (\text{error estándar residual} / \text{la media del consumo sanitario privado de pago})\%$. Así, si el valor del coeficiente de variación es inferior al 10% nos indica que estamos ante un buen modelo, según este criterio.

En nuestro caso el error estándar residual tiene un valor de 121900000, y el coeficiente de variación de 11,6231 % > 10%. Por lo tanto el modelo no es bueno en base a este criterio.

2.- Tabla de Anova: en este criterio cuantificamos cuánta de la variabilidad contenida en los datos ha conseguido ser explicada por dicho modelo. Un modelo es bueno si la variabilidad explicada es mucha, o lo que es lo mismo, si las diferencias entre los datos y las predicciones según el modelo son pequeñas. Construir la tabla de Anova consiste en descomponer la variabilidad de los datos en la parte que es explicada por el modelo y la parte que se deja sin explicar, es decir, la variabilidad de los residuos, y compararlas y valorarlas estadísticamente para ver si la variabilidad explicada por el modelo ajustado es suficientemente grande. En particular, damos respuesta al contraste:

$$H_0: \beta_1 = \beta_2 = \beta_3 = \beta_4 = \beta_5 = \beta_6 = \beta_7 = \beta_8 = \beta_9 = \beta_{10} = 0$$

H_1 : lo contrario

A este respecto, si obtenemos un p-valor inferior a 0,05 rechazaremos H_0 y el modelo será bueno, mientras que, por el contrario, si el p-valor es superior al 0,05 el modelo será malo; contrastamos si las variables explicativas explican suficientemente bien la variable respuesta a través del modelo lineal propuesto.

El p-valor asociado al contraste de la tabla de ANOVA de nuestro modelo es $6,303 \cdot 10^{-6}$ por lo tanto inferior a 0,05, y rechazamos H_0 . En definitiva, el modelo es bueno según este criterio.

3.- Coeficiente de determinación R^2 : otro estadístico útil para chequear la bondad del ajuste se define como la proporción de la varianza que es explicada por la recta de regresión. Obtenemos un número que nos indicará el valor en tanto por ciento de la variabilidad total de los datos que consigue explicar el modelo de regresión creado a tal efecto. Cuando el valor sea cercano a 1 (100%) implicará que buena parte de la varianza es explicada por la recta de regresión y cuando R^2 tenga un valor cercano a cero significará que prácticamente toda la variabilidad de los datos queda sin explicar por el modelo de regresión.

Compararemos, además, el valor del coeficiente de R^2 con el valor de coeficiente R^2 ajustado para ver si son parecidos, ya que cuando hay varias variables explicativas, el mero hecho de tener muchas variables puede distorsionar el R^2 original, por eso lo comparamos con el ajustado. Si el ajustado fuera muy diferente, tomaríamos este como bueno y no el coeficiente de R^2 original.

Para nuestros datos, observamos un valor del coeficiente $R^2 = 0,9943 > 65\%$, por lo que el modelo es bastante bueno, y explica el 99 % de la variabilidad total de la variable Consumo sanitario privado de pago.

Por otra parte, el valor de coeficiente R^2 ajustado = 0,9849, siendo prácticamente igual al coeficiente de R^2 original. Por lo tanto, damos por válido el valor del coeficiente R^2 original 0,9943, que como hemos indicado anteriormente es muy bueno.

FASE 4: DIAGNÓSTICO DEL MODELO

Después del ajuste de un modelo y habiendo superado las pruebas de bondad de ajuste, a continuación verificaremos si el modelo satisface las hipótesis básicas del modelo de regresión lineal para los errores:

- Normalidad
- Varianza constante (Homocedasticidad)
- Incorrelación
- Media cero

El análisis de los residuos nos permitirá detectar deficiencias en la verificación de estas hipótesis, así como descubrir observaciones anómalas o especialmente influyentes en el ajuste. Una vez encontradas las deficiencias, si existen, deberemos analizar si replanteamos el modelo empleando alguna transformación sobre las variables.

La fase del diagnóstico del modelo suele ser gráfica, aunque también existen varios test estadísticos que mostraremos y resolveremos. Los residuos de un modelo lineal los definimos, como habitualmente se hace en la literatura, como las desviaciones entre las observaciones y los valores ajustados.

1.- Normalidad: construiremos, en primer lugar, un histograma para comparar nuestro modelo con la campana de Gauss, cuanto más se desvíe, menos normal será el modelo. Además, analizaremos el gráfico qq-plot, donde se representan los residuos ordenados versus los cuantiles correspondientes de una normal. Si la normalidad de los residuos es cierta, los puntos han de estar alineados con la diagonal. Desviaciones de la diagonal más o menos severas en las colas o en el centro de la distribución implican desviaciones con respecto a la hipótesis de normalidad.

En nuestro caso particular, Figura 6, el histograma no es exactamente una campana de Gauss, desviándose de la simetría y mostrando más peso en la cola de la izquierda de la distribución que en la cola de la derecha.

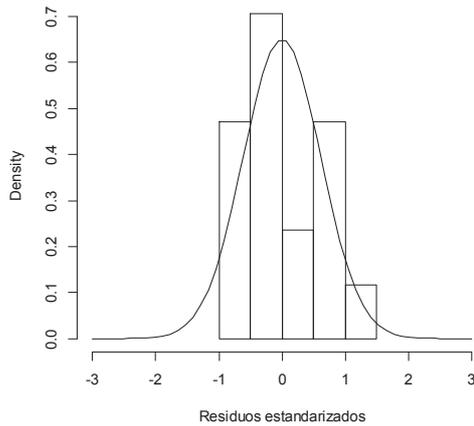


Figura 6. Histograma. Modelo Completo.

Por otra parte, en el gráfico qq-plot la mayoría de los puntos no están sobre la línea discontinua de referencia. Se observan desviaciones de la normalidad en ambas colas, por tanto, existen ciertas dudas sobre la normalidad.

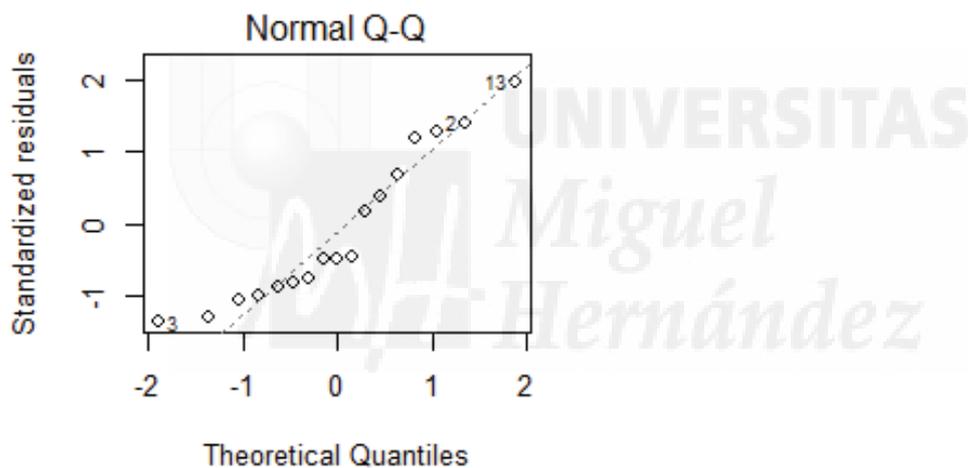


Figura 7. Gráfico QQ-plot. Modelo Completo.

La normalidad también será contrastada a través del contraste de hipótesis de Shapiro-Wilks:

H_0 : $e \sim \text{Normal}$

H_1 : lo contrario

En este caso, obtenemos un p-valor de $0,4418 > 0,05$, por lo que no podemos rechazar la hipótesis de normalidad. En definitiva, esta hipótesis podemos asumir que se verifica. En cualquier caso, nos parece adecuado indicar que el test de normalidad fue aplicado a una muestra de datos de tamaño reducido, por lo que la fiabilidad del propio test estaría en entredicho.

2.- Homocedasticidad (Hipótesis de Varianza Constante): Construiremos un gráfico de residuos versus valores predichos de la variable respuesta Consumo sanitario privado de pago, y numéricamente resolveremos el contraste de hipótesis de Breusch-Pagan.

En el gráfico lo deseable es que los residuos aparezcan representados en una banda horizontal sin tendencias alrededor del cero. Sin embargo, en la Figura 8 vemos que la distribución de los puntos no tiene a lo largo de la gráfica la misma anchura, lo que nos indica que puede haber un problema respecto a la varianza constante. En concreto, la varianza de los errores disminuye con el valor estimado del consumo sanitario privado de pago.

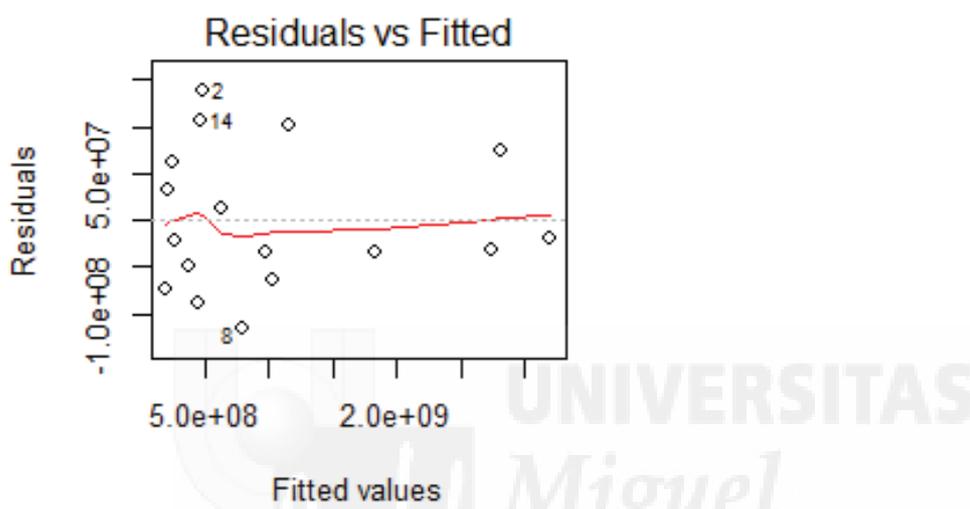


Figura 8. Residuos frente a valores predichos. Modelo Completo.

Resolveremos a continuación el contraste de hipótesis debido a Breusch y Pagan:

H_0 : Homocedasticidad

H_1 : Heterocedasticidad

Obtenemos en concreto un p-valor de $0,9662 > 0,05$. Por lo tanto, hay evidencias estadísticas para no rechazar H_0 . Por lo que asumiremos que la varianza es constante y esta hipótesis se verificaría.

3.- Incorrelación: la incorrelación significa que las observaciones de la variable respuesta no están correlacionadas entre sí, es decir, los valores de unas no afectan a los de otras. A continuación intentaremos comprobar la hipótesis de forma numérica a través de un p-valor. En este sentido, el test que debemos usar se debe a Durbin y Watson:

H_0 : $\rho = 0$ (incorrelación)

H_1 : $\rho \neq 0$

En este caso, obtenemos un p-valor de $0,289 > 0,05$, lo que indica que el modelo es bueno para este criterio y asumiremos que sí se verifica la hipótesis de incorrelación.

4.- La media de los errores es igual a cero: Esta hipótesis siempre se cumple porque usamos el método de máxima verosimilitud, que nos asegura que la media de los residuos siempre es cero.

4.2 PROCEDIMIENTO SECUENCIAL DE SELECCIÓN DE VARIABLES

Mediante el criterio AIC y el método Stepwise vamos a estudiar si nuestro modelo original se puede reducir en número de variables explicativas. El criterio AIC nos indica que valores pequeños son preferidos a valores grandes. Este método nos proporciona un valor AIC global (del modelo completo) y unos valores AIC para el caso de eliminar una cualquiera de nuestras variables explicativas. A continuación, detallamos los resultados obtenidos.

AIC Origianl = 637,32	
VARIABLE	AIC
Población	635,32
Abandono estudios	635,52
Hospitales privados	635,78
Tasa paro	635,87
PIB pc	638,91
Presión fiscal pc	639,63
Esperanza vida	642,73
Mayores 65	643,09
IDH	643,58
Formación superior	650,35

Tabla 6. Stepwise, paso 1.

AIC Origianl = 635,32	
VARIABLE	AIC
Abandono estudios	633,55
Hospitales privados	634,02
Tasa paro	634,64
PIB pc	637,44
Presión fiscal pc	640,42
Esperanza vida	640,74
IDH	646,06
Formación superior	653,10
Mayores 65	681,56

Tabla 7. Stepwise, paso

AIC Original = 633,55	
VARIABLE	AIC
Hospitales privados	632,20
Tasa paro	633,22
PIB pc	635,55
Presión fiscal pc	639,04
Esperanza vida	640,38
IDH	649,46
Formación superior	651,10
Mayores 65	679,71

Tabla 8. Stepwise, paso 3

AIC Original = 632,20	
VARIABLE	AIC
Tasa paro	632,73
PIB pc	635,87
Esperanza vida	639,23
Presión fiscal pc	642,39
IDH	647,60
Formación superior	649,19
Mayores 65	700,22

Tabla 9. Stepwise, paso 4

El criterio AIC ya ha excluido todas las variables menos relevantes y el nuevo modelo resultante es de la siguiente manera:

$$\text{Consumo sanidad privada de pago} = \beta_0 + \beta_1 * \text{Mayores 65} + \beta_2 * \text{PIB pc} + \beta_3 * \text{Tasa paro} + \beta_4 * \text{Presión fiscal pc} + \beta_5 * \text{Formación superior} + \beta_6 * \text{Esperanza vida} + \beta_7 * \text{IDH}$$

4.3 ESTUDIO ECONÓMTRICO DEL MODELO SIMPLIFICADO

Igual que en el modelo completo de regresión múltiple se van a analizar las cuatro fases:

1. Análisis preliminar.
2. Ajuste del modelo.
3. Bondad del ajuste.
4. Diagnóstico del modelo.

FASE 1: ANÁLISIS PRELIMINAR

Como en el modelo anterior vamos a llevar a cabo la inspección gráfica de los datos de la base de datos, para este caso reducida a las variables explicativas mayores 65 años, PIB pc, Tasa de paro, Presión fiscal pc, Formación superior, esperanza de vida e IDH para observar de qué tipo son las relaciones entre las variables si las hay mediante el gráfico de dispersión.

La Figura 9 nos muestra los gráficos de dispersión del modelo reducido y se puede observar de nuevo una relación lineal entre Consumo sanitario privado de pago y mayores de 65 años. Por el contrario, no se aprecia relación lineal con el resto de covariables.

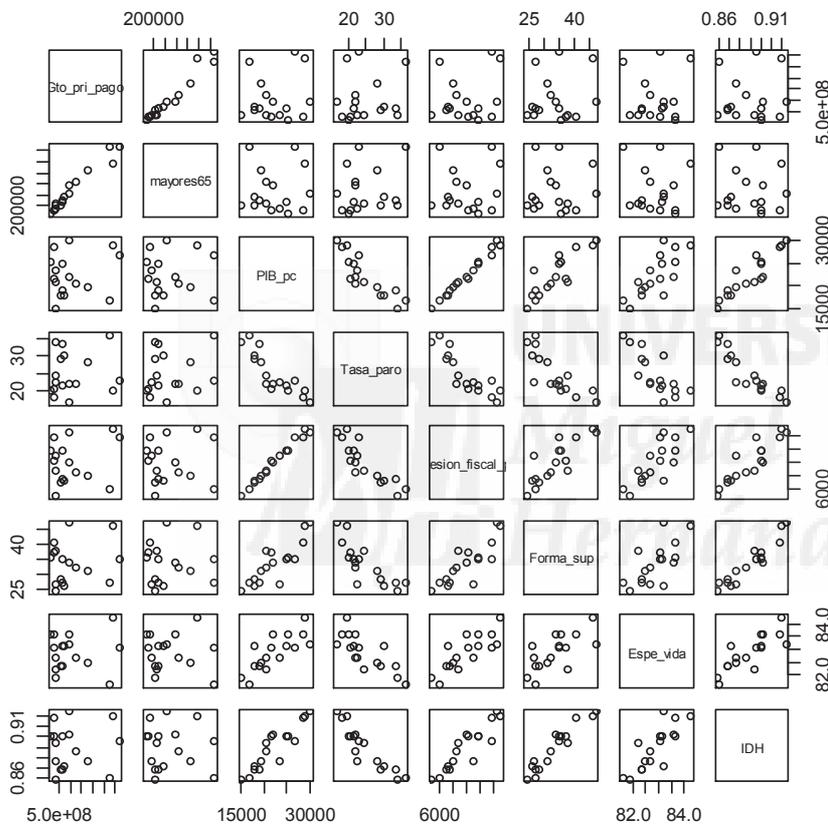


Figura 9. Relación entre la variable respuesta y las variables explicativas. Modelo simplificado.

Realizamos a continuación el análisis numérico para hallar las correlaciones simples y parciales con el fin de corroborar si realmente existen las asociaciones lineales apreciadas en el gráfico.

En la Tabla 10 se aprecia una relación lineal existente entre la variable respuesta Consumo sanitario privado de pago y la variable explicativa mayores de 65 años igual que ocurría en el modelo completo. El resto de variables explicativas no tienen ningún tipo de relación lineal al tener un valor próximo a 0 y por tanto indica la poca influencia sobre el consumo sanitario privado de pago.

VALORES DE LAS CORRELACIONES SIMPLES	
	Gto privado pago
Gto privado pago	1,0000
mayores 65	0,9750
PIB pc	0,1662
Tasa paro	0,1270
Presion fiscal pc	0,2153
Formación superior	0,1679
Esperanza vida	0,0490
IDH	0,0320

Tabla 10. Valores correlación simple. Modelo simplificado.

En la Tabla 11 tenemos representados los valores de las correlaciones parciales, como cabía esperar mayores de 65 años es la que mayor grado de relación lineal tiene con la variable respuesta Consumo sanitario privado de pago con un valor de (0,9918).

VALORES DE LAS CORRELACIONES PARCIALES	
	Gto privado pago
Gto privado pago	0,0000
mayores 65	0,9918
PIB pc	-0,5327
Tasa paro	0,3722
Presion fiscal pc	0,7155
Formación superior	0,8203
Esperanza vida	0,6419
IDH	-0,8005

Tabla 11. Valores correlación parcial. Modelo simplificado.

Cabe destacar como la variable Formación superior cuando se relaciona con las demás variables tiene un peso muy importante en nuestro modelo simplificado.

FASE 2: AJUSTE DEL MODELO

En la segunda fase vemos los valores que estima el modelo para β_1 , β_2 , β_3 , β_4 , β_5 , β_6 y β_7 , sus respectivos intervalos de confianza al 95% y el contraste de hipótesis para responder a la cuestión de si cada coeficiente es igual a cero o no. Por lo tanto cuando rechazamos la hipótesis de coeficiente cero obtendremos las variables que son útiles, es decir, descubriremos las variables que contienen suficiente información para explicar la variable consumo sanitario privado de pago. Estimaremos el valor de los parámetros con el criterio de máxima verosimilitud.

En la Tabla 12 mostramos las estimaciones máximo verosímiles de los parámetros, así como el p-valor asociado a los contrastes anteriormente explicados. Como en el modelo completo añadimos valores relacionados con la bondad del ajuste.

El modelo ajustado para el estudio reducido corresponde a la expresión siguiente:

$$\text{Consumo sanidad privada de pago} = 7,814 \cdot 10^9 + 2,091 \cdot 10^3 \cdot \text{Mayores 65} - 1,055 \cdot 10^5 \cdot \text{PIB pc} + 1,634 \cdot 10^7 \cdot \text{Tasa paro} + 5,220 \cdot 10^5 \cdot \text{Presión fiscal pc} + 7,010 \cdot 10^7 \cdot \text{Formación superior} + 2,085 \cdot 10^8 \cdot \text{Esperanza vida} - 3,331 \cdot 10^{10} \cdot \text{IDH}$$

RESUMEN DEL MODELO								
	β_0	β_1	β_2	β_3	β_4	β_5	β_6	β_7
Parámetro	$7,814 \cdot 10^9$	$2,091 \cdot 10^3$	$(-), 1,055 \cdot 10^5$	$1,634 \cdot 10^7$	$5,220 \cdot 10^5$	$7,010 \cdot 10^7$	$2,085 \cdot 10^8$	$(-), 3,331 \cdot 10^{10}$
p-valor	0,255	$2,32 \cdot 10^{-9}$	0,091	0,259	0,013	0,001	0,0332	0,003
Residual Standard Error	102100000							
CV	9,735							
Multiple R-squared	0,994							
Ajusted R-squared	0,989							
p-valor	$2,828 \cdot 10^{-9}$							

Tabla 12. Resumen Modelo simplificado.

En la Tabla 13 mostramos los intervalos de confianza al 95% para cada uno de los coeficientes del modelo lineal.

INTERVALO DE CONFIANZA		
	Extremo inferior	Extremo superior
Intercepto	$-6,744215 \cdot 10^9$	$2,237292 \cdot 10^{10}$
Mayores 65	$1,888599 \cdot 10^3$	$2,294053 \cdot 10^3$
PIB pc	$-2,319527 \cdot 10^5$	$2,087050 \cdot 10^4$
Tasa paro	$-1,438633 \cdot 10^7$	$4,706961 \cdot 10^7$
Presión fiscal pc	$1,377657 \cdot 10^5$	$9,062162 \cdot 10^6$
Forma superior	$3,325039 \cdot 10^7$	$1,069531 \cdot 10^8$
Esperanza vida	$2,072031 \cdot 10^7$	$3,962742 \cdot 10^8$
IDH	$-5,211143 \cdot 10^{10}$	$-1,450343 \cdot 10^{10}$

Tabla 13. Intervalos de confianza. Modelo simplificado.

$$IC_{95\%} \beta_1 = (1,888599 \cdot 10^3, 2,294053 \cdot 10^3)$$

$$IC_{95\%} \beta_2 = (-2,319527 \cdot 10^5, 2,087050 \cdot 10^4)$$

$$IC_{95\%} \beta_3 = (-1,438633 \cdot 10^7, 4,706961 \cdot 10^7)$$

$$IC_{95\%} \beta_4 = (1,377657 \cdot 10^5, 9,062162 \cdot 10^6)$$

$$IC_{95\%} \beta_5 = (3,325039 \cdot 10^7, 1,069531 \cdot 10^8)$$

$$IC_{95\%} \beta_6 = (2,072031 \cdot 10^7, 3,962742 \cdot 10^8)$$

$$IC_{95\%} \beta_7 = (-5,211143 \cdot 10^{10}, -1,450343 \cdot 10^{10})$$

Realizamos el contraste de hipótesis para comprobar si los coeficientes contienen la información suficiente para explicar la variable explicativa. En la Tabla 12 tenemos los p-valores asociados para cada par de coeficiente-contraste.

Para el caso de beta 1, el contraste a resolver sería el siguiente:

$$H_0: \beta_1 = 0$$

$$H_1: \beta_1 \neq 0$$

El p-valor de este par de contraste-coeficiente es igual a $2,32 \cdot 10^{-9} < 0,05$ por tanto rechazamos H_0 y asumimos que el beta 1 distinto de cero.

Para el resto de variables obtenemos los siguientes valores:

p-valor $\beta_2 = 0,091 > 0,05$ No rechazamos H_0 , asumiremos que el parámetro es nulo.

p-valor $\beta_3 = 0,259 > 0,05$ No rechazamos H_0 , asumiremos que el parámetro es nulo.

p-valor $\beta_4 = 0,013 > 0,05$ Rechazamos H_0 , el parámetro es distinto de cero.

p-valor $\beta_5 = 0,001 > 0,05$ Rechazamos H_0 , el parámetro es distinto de cero.

p-valor $\beta_6 = 0,033 > 0,05$ Rechazamos H_0 , el parámetro es distinto de cero.

p-valor $\beta_7 = 0,003 < 0,05$ Rechazamos H_0 , el parámetro es distinto de cero.

Con la información obtenida en este ajuste podemos apreciar todos los p-valor a excepción de β_2 y β_3 son inferiores a 0,05. Por lo tanto según el método de contraste de hipótesis hemos mejorado considerablemente el modelo logrando un buen modelo.

FASE 3: BONDAD DEL AJUSTE

Al igual que cuando realizamos el ajuste en un modelo de regresión lineal del modelo completo, en el nuevo modelo ajustado debemos de verificar si dicho modelo proporciona un buen ajuste a la hora de predecir la variable respuesta. Utilizaremos los mismos tres modos utilizados para el modelo completo.

1.- Error estándar residual: hacemos la medición de la bondad del ajuste relativa a la escala de medida utilizada. De los datos de la Tabla 14 obtenemos el valor residual estándar del modelo, el cual tiene un valor de 102.100.000. Para determinar si este valor es grande o pequeño realizamos el cálculo del coeficiente de variación $cv = 100 \cdot (102.100.000 / \text{la media del consumo sanitario privado de pago}) \%$ lo que da un resultado de $cv = 9,735$ que al ser menor del 10% nos indica que el modelo es bueno. El valor es menor que en modelo global que era 11,6231 por tanto el nuevo modelo mejora el criterio respecto al modelo anterior.

2.- Tabla de Anova: en este criterio cuantificamos cuánta de la variabilidad contenida en los datos ha conseguido ser explicada por dicho modelo.

Damos respuesta al siguiente contraste de hipótesis:

$H_0: \beta_1 = \beta_2 = \beta_3 = \beta_4 = \beta_5 = \beta_6 = \beta_7 = 0$

H_1 : lo contrario

El p-valor de nuestro modelo es $2,828 \cdot 10^{-9}$ que como es menor que 0,05, rechazo H_0 y asumimos que el modelo es bueno al igual que en el modelo completo. Las variables explicativas explican suficientemente bien la variable respuesta a través del modelo lineal propuesto.

3.- Coeficiente de determinación: obtenemos un valor que nos indicará el valor en tanto por ciento de la variabilidad total de los datos, será la proporción de la varianza que es explicada por la recta de regresión. Cuanto más cercano a 1 (100%) sea el valor, más parte de la varianza es explicada por la recta de regresión, y cuánto más se acerque a cero el valor, indica que prácticamente toda la variabilidad de los datos queda sin explicar por el modelo de regresión.

Como anteriormente compararemos el valor del coeficiente de R^2 con el valor de coeficiente R^2 ajustado para ver si son parecidos, en caso de que no lo sean descartaremos el valor del coeficiente de R^2 por estar distorsionado y daremos por válido el coeficiente de R^2 ajustado.

El coeficiente de R^2 es 0,994, y por tanto la recta de regresión explica con bastante precisión la variabilidad total de los datos, siendo un valor muy alto, por lo que es un modelo bastante bueno.

El coeficiente ajustado de R^2 tiene un valor similar de 0,989 por lo que consideramos bueno el valor de coeficiente de R^2 . Anteriormente en el modelo completo tenía un valor del 0,994, por lo que este modelo simplificado sigue explicando bien la variabilidad de la variable respuesta.

FASE 4: DIAGNÓSTICO DEL MODELO

Habiendo ajustado el modelo y superadas las pruebas de bondad del ajuste en esta fase vamos a verificar si satisface las hipótesis básicas del modelo de regresión lineal para los errores.

Mediante este análisis detectaremos deficiencias en la verificación de las hipótesis. De momento no nos planteamos la transformación de alguna de las variables al superar hasta ahora todos los test realizados, no obstante, si no se cumple alguna de las siguientes hipótesis, normalidad, homocedasticidad, varianza constante e incorrelación, tendremos que rechazar el modelo simplificado y nos replantearemos un modelo empleando alguna transformación sobre las variables.

Para estudiar las desviaciones entre las desviaciones entre las observaciones y los valores ajustados haremos a continuación el diagnóstico gráfico y numérico a través de diferentes test estadísticos.

1.- Normalidad: haremos un estudio gráfico viendo el histograma y el gráfico QQPLOT donde representamos los residuos ordenados versus los cuantiles correspondientes de una normal como en el modelo anterior.

Vemos que el histograma, Figura 10. Las colas de los valores ajustados no siguen la tendencia de la normalidad de la campana de Gauss y el histograma también representa una clara desviación de la simetría mostrando más peso en la cola de la izquierda de la distribución tal y como hacía en el modelo completo.

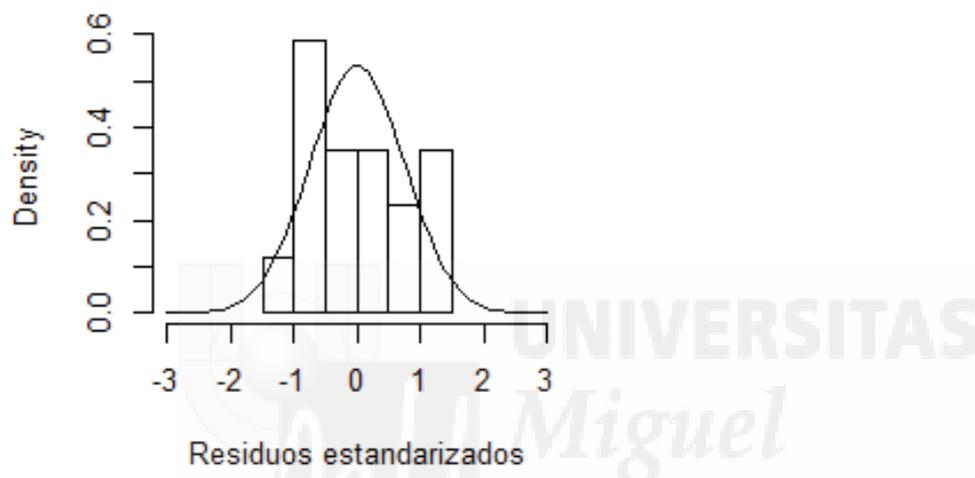


Figura 10. Histograma. Modelo simplificado.

Por otra parte, el gráfico qqplot, Figura 11. Tal y como ocurría en el modelo completo la mayoría de los puntos no están sobre la línea discontinua de referencia. Se observan desviaciones de la normalidad en ambas colas, por tanto, existen ciertas dudas sobre la normalidad.

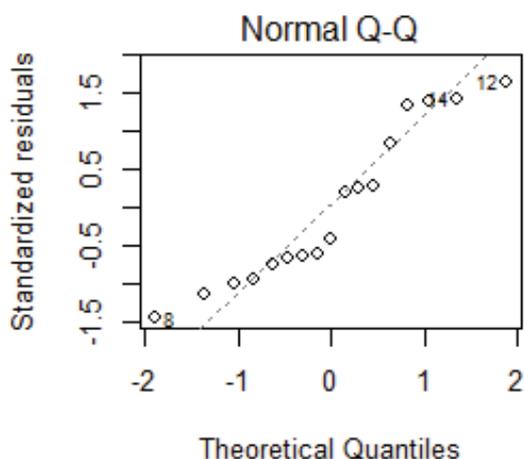


Figura 11. Gráfico QQ-plot. Modelo simplificado.

Para despejar las dudas existentes sobre la normalidad de las pruebas anteriores hacemos una comprobación numérica mediante el contraste de hipótesis de Shapiro-Wilks:

H_0 : $e \sim \text{Normal}$

H_1 : lo contrario

El valor obtenido es un p-valor de $0,0989 > 0,05$ por lo tanto, asumimos que sí se verifica la hipótesis de normalidad.

2.- Homocedasticidad: (hipótesis de varianza constante): Realizaremos un análisis sobre el gráfico de residuos versus valores predichos de la variable respuesta consumo sanitario privado de pago y numéricamente resolveremos el contraste de hipótesis de Breusch-Pagan.

En el gráfico de la Figura 12. Se aprecia que la distribución de los puntos no tiene a lo largo de la gráfica la misma anchura, a la izquierda es mucho más ancho, estrechándose por el centro y volviéndose a ensanchar por la derecha. Lo ideal es que fuese una banda lo más uniforme posible respecto del eje de abscisas, por tanto, existen dudas sobre el cumplimiento de la homocedasticidad.

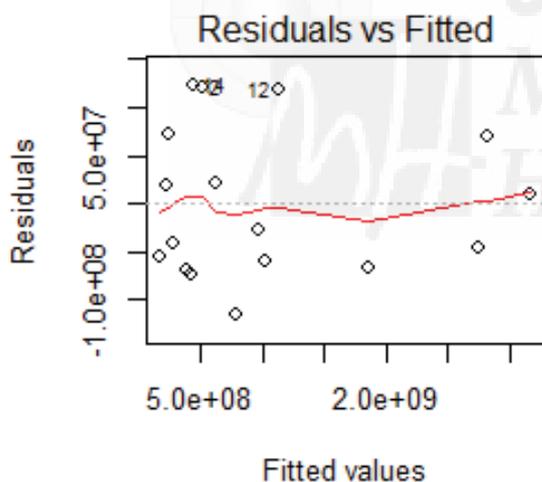


Figura 12. Residuos frente a valores predichos. Modelo simplificado.

Para asegurarnos numéricamente del cumplimiento de ésta hipótesis realizamos el contraste de hipótesis de Breusch-Pagan:

H_0 : Homocedasticidad

H_1 : Heterocedasticidad

Obtenemos un p-valor de $0,586 > 0,05$, por lo que no podemos rechazar H_0 lo que implica que asumimos que el modelo cumple con la hipótesis de varianza constante, pese a la primera observación gráfica.

3.- Incorrección: Para establecer si los valores de unas variables afectan a las de otras realizamos el contraste de hipótesis de Durbin y Watson.

$$H_0: \rho = 0$$

$$H_1: \rho \neq 0$$

Se observa un p-valor de $0,4531 > 0,05$ que indica que el modelo es bueno y asumimos que cumple el criterio de Incorrección.

4.-Media cero: siempre se cumple porque usamos el modelo de máxima verosimilitud y está implícito que la media de los errores es siempre cero.

Al igual que con el modelo completo, el modelo simplificado cumple con las bondades del ajuste y se verifican las hipótesis asumidas.

4.4 TRANSFORMACIÓN DE LA VARIABLE RESPUESTA

Pese a verificarse las hipótesis asumidas por el modelo ajustado, vamos a realizar una transformación de la variable respuesta con el fin de observar si mejoran sobre todo los gráficos del diagnóstico de los residuos.

Para determinar qué transformación es aceptable nos apoyamos inicialmente en un método de transformación que pertenece a la familia de transformaciones potenciales, como es el caso de las transformaciones Box-Cox. Este método se basa en el resultado que arroja el valor lambda, cuando éste es distinto de cero es recomendable que la transformación sea la variable respuesta elevada al valor lambda, y cuando lambda valga cero la transformación recomendable es el logaritmo de la variable respuesta. La transformación de Box-Cox es aplicable sólo cuando la respuesta es estrictamente positiva. Si hay valores cero o negativos, la forma habitual de proceder es añadiendo una constante a la respuesta antes de aplicar el método, con el fin de hacer todos los valores positivos; sin embargo, suele haber poca información en los datos para elegir dicha constante.

Hemos obtenido mediante la transformación de Box-Cox un valor de lambda de 1 lo que nos indica que el modelo original es muy bueno. Aún así, realizamos la transformación logarítmica de la variable respuesta para intentar mejorar el modelo. Al observar la transformación de la variable respuesta en términos numéricos vemos que la transformación ha reducido en uno todas las variables respuestas. Por lo tanto, desestimamos la realización del estudio econométrico de la variable transformada ya que el resultado sería idéntico al del modelo simplificado.

5. CONCLUSIONES

A lo largo del presente estudio hemos analizado tres modelos diferentes pero estrechamente relacionados, un primer análisis con el modelo lineal completo, a continuación, un segundo análisis simplificado en el que hemos eliminado del modelo

las variables explicativas que no aportaban información relevante y podían ser excluidas del mismo, y por último, el modelo con la variable respuesta “Consumo sanitario privado de pago” transformada. Todo ello para averiguar si existe algún tipo de relación lineal entre la variable respuesta y las variables explicativas descritas anteriormente.

El procedimiento empleado para el análisis del primer y segundo modelo ha sido el mismo, primero hemos realizado un análisis preliminar, a continuación procedemos a ajustar el modelo, comprobamos la bondad de los ajustes y por último verificamos que se cumplen las hipótesis asumidas por el modelo. El tercer modelo no hemos llegado a realizarlo debido a que en la transformación lambda nos ha dado 1 y al ser mínima el resultado hubiera sido el mismo que en el modelo simplificado.

En el análisis preliminar, en primer lugar, hemos analizado gráfica y numéricamente las relaciones existentes entre las variables explicativas y nuestra variable respuesta. Gráficamente los gráficos de dispersión nos han mostrado una clara relación lineal entre la variable respuesta Consumo sanitario de pago y las variables explicativas población, número de hospitales privados y mayores de 65 años. En el resto de variables explicativas no se aprecia relación lineal. Con el estudio de las correlaciones simples corroboramos lo observado gráficamente con las variables explicativas, es decir, una fuerte relación lineal de las variables mencionadas anteriormente. Sin embargo, con el estudio de las correlaciones parciales, destaca el hecho que la variable Formación superior, que en un principio parecía no tener importancia en el modelo, al relacionarla con las demás variables es la de mayor importancia, lo que nos indica que la información contenida es relevante para el modelo.

Finalizado el análisis preliminar procedemos a ajustar el modelo para estimar los valores de los diez parámetros del modelo completo y los siete parámetros del modelo simplificado y transformado, además calculamos sus intervalos de confianza al 95%. En el primer modelo asumimos que los parámetros pueden ser nulos excepto el coeficiente de Formación superior, siendo esta variable la que contiene información suficientemente útil para explicar la variable respuesta. En los dos modelos restantes, simplificado y transformado, de las siete variables explicativas, en cinco de ellas (Mayores de 65 años, Presión fiscal pc, Formación superior, Esperanza de vida e IDH) rechazamos la hipótesis de $H_0: \beta_1 = 0$ y por tanto los parámetros son distintos de cero. Con lo cual, la información contenida en dichas variables sí es útil para explicar los modelos. Con las variables PIB pc y Tasa de paro no podemos rechazar la hipótesis, asumimos que el parámetro es nulo y la información de la variable no es estadísticamente significativa para explicar la variable respuesta, a un nivel de error del 5%.

Posteriormente hemos realizado la bondad del ajuste, y así poder verificar si nuestro modelo es bueno para explicar la variable respuesta. Para ello hemos utilizado tres criterios: Error Estándar Residual, la Tabla de ANOVA y del Coeficiente de Determinación. En los tres modelos analizados los tres criterios anteriores han sido favorables y por tanto nos indican que son buenos.

Por último, hacemos un diagnóstico de los modelos y comprobamos si éstos cumplen las hipótesis básicas de regresión lineal para los errores. Para ello nos hemos ayudado tanto de gráficas (histograma y gráfico qqplot en la normalidad y residuos frente a valores predichos en la homocedasticidad). Además se utiliza el contraste de hipótesis (Shapiro-Wilks en la normalidad, Breusch-Pagan en la homocedasticidad y Durbin-Watson para la incorrelación). Por otra parte, la hipótesis de media cero se cumple siempre por usar el método de máxima verosimilitud. En los modelos las pruebas gráficas parecen indicar que no cumplen las hipótesis básicas, sin embargo, al realizar los contrastes de hipótesis, éstas se cumplen y por tanto podemos aceptar que los modelos son buenos.

En resumen, desde un principio y con la salvedad de que nuestra base de datos es reducida debido a la peculiaridad de las muestras, nos indica que los modelos son buenos y existe cierta relación lineal entre las variables explicativas y el importe de consumo sanitario privado de pago. Con el modelo simplificado tratamos de reducir y eliminar las variables que no son útiles para hacer un modelo más sencillo y también ha cumplido los objetivos previstos. El motivo de hacer una transformación de la variable respuesta ha sido para comprobar si todavía podíamos mejorar el modelo simplificado, sobre todo en los análisis gráficos. Una vez transformadas las variables respuesta hemos observado que no se podía mejorar ya que la transformación ha sido prácticamente insignificante. Únicamente ha reducido las variables respuesta en una unidad.

Finalizamos la realización de este análisis con la conclusión de que a mayor actividad en el sector sanitario privado en una Comunidad Autónoma mayor consumo sanitario privado de pago y por lo tanto mercados más atractivos. Consideramos en base al presente estudio que los modelos son buenos y es importante resaltar el comportamiento de las variables. Por ejemplo, a mayor población con más de 65 años, mayor esperanza de vida, mayor presión fiscal y mayor formación superior de los individuos, mayor consumo sanitario privado de pago. Las dos primeras variables comentadas están relacionadas con que haya más población mayor que es en los años donde más sanidad se consume. Las dos segundas variables están muy relacionadas con mayores salarios y por lo tanto mayor capacidad de consumo. Por el contrario, menor valoración en IDH, mayor consumo sanitario privado de pago manteniendo el resto de variables constantes. Esto se debe a que las comunidades con mayor IDH tienen mejor sanidad pública, por lo que menos individuos terminan acudiendo a la sanidad privada.

6. BIBLIOGRAFÍA

Aparicio, J. (s.f.). *ECONOMETRÍA*.

<http://umh3067.edu.umh.es/material/videos/>

Aparicio, J., Martínez Mayoral, M., & Morales, J. (s.f.). *Modelos Lineales Aplicados en R*. (C. I. Hernández, Ed.) Dto. Estadística, Matemáticas e Informática.

Berechet, C. (2013) Día de la liberación fiscal. 10 de mayo de 2013. Última fecha de acceso: 14/06/2015.

<http://www.civismo.org/es/investigaciones/informes/dia-de-la-liberacion-fiscal-2013>

Fundación IDIS. (2014). *Sanidad privada, aportando valor. Análisis situación 2014*.

Última fecha de acceso: 20/06/2015.

<https://www.fundacionidis.com/es/informes/informe-sanidad-privada-aportando-valor-an%C3%A1lisis-de-situaci%C3%B3n-2014>

Instituto Nacional de Estadística

<https://www.ine.es/>

Sánchez Bayle, M., & Martín García, M. (2013). *Situación, amenazas y estrategias para defender la sanidad pública*. Última fecha de acceso: 21/06/2015.

<http://www.nuevatribuna.es/articulo/sanidad/situacion-amenazas-y-estrategias-defender-sanidad-publica/20131111172701098173.html>

Ministerio de Sanidad, Asuntos Sociales e Igualdad. (2013). *Catálogo Nacional de Hospitales 2013*.

PwC, equipo de sanidad y farmacia. (2012). *Diez temas candentes de la Sanidad española para 2013*. Última fecha de acceso: 20/06/2015.

<http://www.pwc.es/es/publicaciones/sector-publico/diez-temas-candentes-sanidad-2013.jhtml>