

DETERMINACIÓN DE LA RELACIÓN ENTRE EL TAMAÑO DEL CAPITAL HUMANO DE  
UN CONJUNTO DE AEROLÍNEAS INTERNACIONALES Y UNA SERIE DE VARIABLES DE  
OPERACIÓN DE LAS EMPRESAS.

**TRABAJO FIN DE GRADO**

**CURSO 2014-2015**



Autor: Rosana García Garre

Tutor: Juan Aparicio Baeza

Fecha de entrega: 4 de junio de 2015

Grado de Administración y Dirección de Empresa

Facultad Ciencias Sociales y Jurídicas de Orihuela

Universidad Miguel Hernández de Elche

## INDICE

RESUMEN .....	3
<b>1. INTRODUCCIÓN .....</b>	<b>4</b>
<b>1.1 El segmento de las aerolíneas.....</b>	<b>5</b>
<b>2. OBJETIVOS.....</b>	<b>15</b>
<b>3. METODOLOGÍA .....</b>	<b>15</b>
<b>3.1 Muestra de datos .....</b>	<b>16</b>
<b>4. ESTUDIO ECONÓMETRICO .....</b>	<b>17</b>
<b>4.1 ESTUDIO ECONÓMETRICO DEL MODELO COMPLETO.....</b>	<b>17</b>
FASE 1: ANÁLISIS PRELIMINAR.....	18
FASE 2: AJUSTE DEL MODELO.....	21
FASE 3: BONDAD DEL AJUSTE.....	23
FASE 4: DIAGNÓSTICO DEL MODELO .....	25
<b>4.2-MÉTODO DE SIGNIFICATIVIDAD ESTADÍSTICA .....</b>	<b>31</b>
<b>4.3-ESTUDIO ECONÓMETRICO DEL MODELO SIMPLIFICADO .....</b>	<b>32</b>
FASE 1: ANÁLISIS PRELIMINAR.....	32
FASE 2: AJUSTE DEL MODELO .....	34
FASE 3: BONDAD DEL AJUSTE.....	35
FASE 4: DIAGNÓSTICO DEL MODELO .....	36
<b>4.4-TRANSFORMACIÓN DE LA VARIABLE RESPUESTA.....</b>	<b>40</b>
FASE 1: ANÁLISIS PRELIMINAR.....	41
FASE 2: AJUSTE DEL MODELO .....	42
FASE 3: BONDAD DEL AJUSTE.....	43
FASE 4: DIAGNÓSTICO DEL MODELO .....	44
<b>6. CONCLUSIONES .....</b>	<b>48</b>
BIBLIOGRAFIA .....	51

## RESUMEN

Este estudio se realiza con el fin de determinar si existe algún tipo de relación estadística, entre el número de empleados de las aerolíneas y una serie de variables económicas relacionadas con estas compañías.

El año utilizado de referencia para este estudio es el 2009. Los datos hacen referencia a las aerolíneas del mundo. Con todos ellos hemos elaborado una base de datos que nos ha permitido obtener la información oportuna y necesaria para poder desarrollar nuestro trabajo.

En la introducción hemos hablado de la importancia del transporte aéreo para el desarrollo económico de los países y como este ha hecho cambiar la estructura y estrategias del sector aeroportuario, dando lugar a la aparición de compañías de bajo coste donde el control de los gastos es fundamental para la rentabilidad de las mismas.

El estudio econométrico está compuesto de tres estudios con el fin de determinar cual de ellos contribuye a explicar mejor el aumento o disminución del número de empleados de las aerolíneas. El primer modelo, incluye todas las variables, después hemos eliminado aquellas que no son relevantes para el modelo y finalmente hemos transformado la variable respuesta para comprobar si se podían mejorar los resultados del modelo anterior.

Finalmente, hemos comprobado que en todos los modelos analizados destaca la relación lineal entre la variable respuesta y la variable explicativa Passenger (número medio de kilómetros que cada pasajero ha sido transportado). En el modelo completo, otras variables pese a tener una relación lineal positiva, es poco importante y por tanto no contribuyen a explicar la variable respuesta. Esto ha dado lugar a que en el modelo reducido, excepto la variable explicativa Passenger, todas las demás son eliminadas por carecer de valor explicativo. De ahí que podamos afirmar que si el número medio de kilómetros que cada pasajero ha sido transportado aumenta en 1000 unidades (kilómetros) el personal de plantilla se ve incrementado en 5381 unidades.

## 1. INTRODUCCIÓN

El transporte aéreo, tanto de pasajeros como de mercancías, es uno de los sectores de mayor importancia en las economías desarrolladas desde comienzos del siglo XX. Sus fuertes vínculos con el comercio, el turismo y los negocios han dado lugar a que el sector presente tasas medias de crecimiento anuales de entre un 5 y un 8% en los países desarrollados. Además, este sector es profundamente dinámico, en particular, desde su liberalización, lo que ha facilitado la aparición de nuevos modelos de negocio y la existencia de una fuerte competencia que ha reducido los precios del transporte aéreo.

Estos cambios en el sector del transporte aéreo han impactado en el sector aeroportuario, elemento estrictamente necesario para su desarrollo. Así, los aeropuertos han pasado de considerarse un monopolio natural a una unidad de negocio potencialmente rentable. De hecho, en términos generales, la evidencia muestra en el sector unas tasas de rentabilidad importantes que en parte explican la tendencia general a la privatización.

Por esta razón, estas infraestructuras compiten por una demanda que la componen, por un lado, los pasajeros, que además de clientes de las aerolíneas también lo son de las actividades comerciales realizadas en el aeropuerto y, por otro, las aerolíneas.

Estas últimas, con la aparición de las compañías de bajo coste (LCC) y la reducción de costes de cambio de un aeropuerto a otro, han incrementado la presión competitiva entre dichos aeropuertos.

Sin embargo, las mayores presiones competitivas que los nuevos modelos de negocio aparecidos en el sector aéreo ejercen sobre los aeropuertos, no impiden que, en determinados casos, alguno de ellos cuente con un elevado poder de mercado que requiera de su regulación.

Dada la evolución anterior, el sector aeroportuario ha observado importantes cambios en los últimos años con la introducción de capital y modelos de gestión privada de los mismos, tanto total como de forma mixta, modelos más flexibles de gestión descentralizada y la aplicación de soluciones regulatorias allí donde se identifica la existencia de aeropuertos con poder de mercado.

De la comparativa internacional se comprueba que el caso español es el único país con un sistema de gestión centralizada y conjunta en manos de un único gestor, con un volumen elevado tanto de infraestructuras como de pasajeros transportados.

A pesar de la situación española se puede comprobar que los aeropuertos (como lo demuestra la evolución de las tasas aeroportuarias así como su composición), muestran una progresiva tendencia a incrementar aquéllas tasas que afectan a los pasajeros en detrimento de las abonadas por las aerolíneas, más sensibles al precio.

### **1.1 El segmento de las aerolíneas**

Podemos considerar, como hemos indicado anteriormente, el segmento de las aerolíneas como demandantes y usuarias de los servicios aeroportuarios.

La evolución de estos, tras los procesos liberalizadores, que de una forma más o menos generalizada tuvieron lugar a partir de la década de los años 70, puede considerarse como uno de los principales impulsores de los cambios sucedidos en la concepción de los aeropuertos.

Dentro de este segmento se puede distinguir a su vez entre dos subsegmentos:

- Subsegmento de las líneas aéreas de bajo coste.
- Subsegmento de las líneas aéreas tradicionales.

El subsegmento de las líneas aéreas tradicionales representa todo el conjunto de las líneas aéreas que no pertenezcan al sector de bajo coste. En este contexto, nos referimos a todas las líneas aéreas previamente existentes, consideradas como nacionales, regionales y chárter que no han efectuado todavía los cambios estratégicos correspondientes, con el fin de transformar su modelo de negocio en un modelo de operación recién desarrollado como es el de bajo coste.

Por otro lado, (Meissner M, 2008) el término “líneas aéreas de bajo coste” procede de la traducción inglesa “Low Cost Airlines”<sup>1</sup>. En la literatura inglesa existen tres definiciones que

se relacionan en este contexto: Low Fare (bajo precio), Low Cost (bajo coste), y No Frills (sin adornos). Estas líneas aéreas se caracterizan porque generalmente ofrece una tarifa

barata a cambio de eliminar muchos de los servicios que tradicionalmente recibían los pasajeros o de cobrar adicionalmente por ellos.

Con referencia a la comparación entre ambas líneas aéreas, en la práctica es más la cuestión de la percepción tanto interna (propia) como externa (ajena).

En el caso de la percepción interna hay que analizar cómo se presenta la compañía aérea a sí misma ante el público<sup>2</sup>.

El resultado de este análisis debe ser homogéneo, es decir, cada compañía aérea debe poder mostrarse de una forma clara a un sector particular del mercado aéreo ante el público.

En el caso de la percepción externa hay que analizar cómo percibe el público a la compañía aérea.

El resultado del segundo análisis puede ser heterogéneo, es decir, los diferentes participantes del público pueden asignar la misma compañía aérea a diferentes sectores de mercado aéreo. De esta forma, existe no solamente heterogeneidad dentro de los resultados de la percepción externa, sino también, como consecuencia, una heterogeneidad de los resultados dentro de la comparación de la percepción interna con la externa. Por tanto, no puede existir una clara homogeneidad del resultado de la comparación de las líneas aéreas con relación a la aplicación de dicha definición en la práctica.

Actualmente, se observa a nivel global una acusada tendencia hacia el desarrollo y consolidación de las denominadas aerolíneas de bajo coste en detrimento de las aerolíneas tradicionales, fenómeno con origen en Estados Unidos pero con a la práctica totalidad del sector internacional.

---

<sup>1</sup>Fuente: Gross, S. y Schröder, A. (2004): *Low Cost Airlines in Europa*. Erich Schmidt Verlag GmbH & Co., Berlin. Pp. 3-4. Pohl, A. (2004): *Chancen und Risiken von No-Frills Airlines und deren Auswirkungen auf den Flughafen Bonn*. Hausarbeit Fachhochschule Bonn-Rhein-Sieg.

<sup>2</sup> Como público se entiende el conjunto del entorno, tanto los consumidores, competidores como otros participantes del mercado.

Estos cambios en los modelos de negocio de las aerolíneas, se han derivado hacia esquemas más flexibles. Las mayores facilidades de movilización de sus bases estratégicas, han tenido una notable influencia en la modificación, a su vez, del propio sector aeroportuario, imprimiendo en cierta forma, restricciones al desarrollo de la política comercial de los aeropuertos.

En términos generales, puede caracterizarse a las aerolíneas de bajo coste como aquéllas centradas en el desarrollo del segmento de vuelos punto a punto<sup>3</sup>, con una extensiva utilización de las aeronaves y mínima utilización del personal de a bordo, con preferencia de uso de aeropuertos secundarios que disponen de una mayor facilidad de acceso a slots o derechos de despegue y aterrizaje y especializadas únicamente en el transporte de pasajeros, abandonando el segmento de mercancías.

Solo utilizan aquellas infraestructuras indispensables para la puesta en marcha de los vuelos, los mínimos tiempos de conexión o los salarios inferiores, a la media del sector, de su personal.

Todo esto tiene implicaciones sobre las tarifas que las mismas posteriormente pueden cobrar a sus usuarios, situándose éstas en niveles inferiores a las de las compañías tradicionales y no permitiendo, por regla general, una discriminación por clases de pasajeros.

---

<sup>3</sup> Los modelos de conexión de las aerolíneas se dividen en modelos punto a punto, que conectan pares de ciudades y los modelos centro-radio (hub and spoke) que une destinos radiales con un aeropuerto centro que sirve de base para acceder posteriormente a otros destinos, permitiendo reducciones de costes y aprovechamientos de economías de escala.

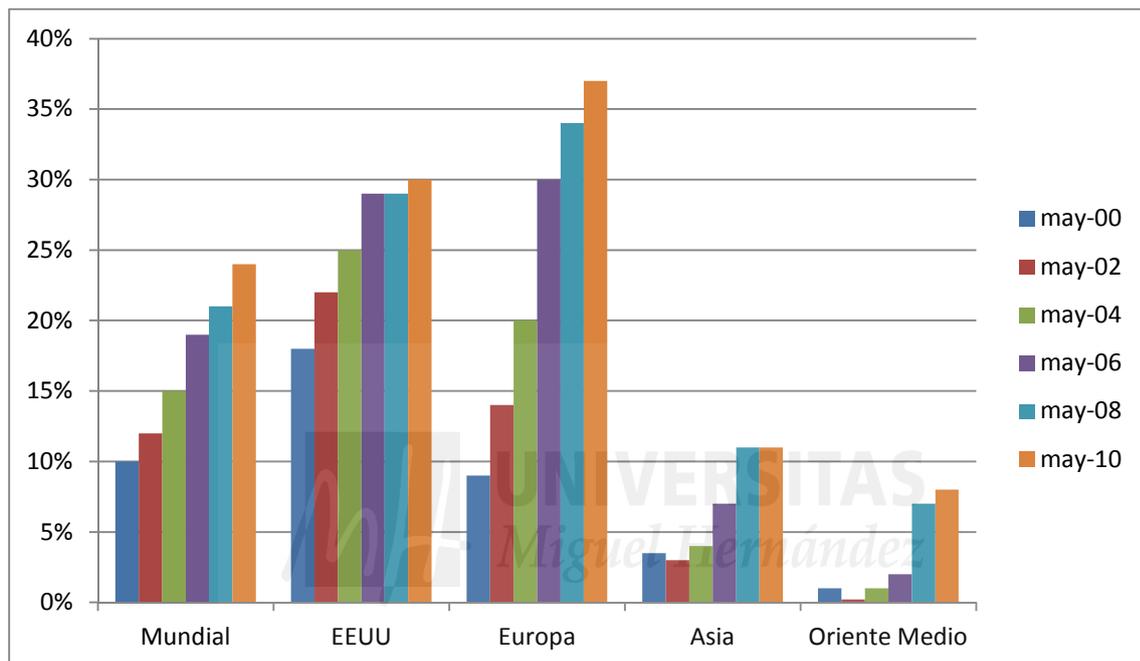
La Tabla 1 recoge las principales diferencias entre las aerolíneas tradicionales y las de bajo coste.

**Tabla 1. Comparación entre aerolíneas de bajo coste y tradicionales**

	<b>AEROLÍNEAS DE BAJO COSTE</b>		<b>AEROLÍNEAS TRADICIONALES</b>
<b>Aeropuertos</b>	Fundamentalmente secundarios. Uso de los servicios indispensables y poco tiempo de estancia.		Principales. Mayores servicios y tiempos de estancia.
<b>Conexiones</b>	De corto radio. Centradas en el punto a punto.		Con posibilidad de escalas y vuelos de interconexión.
<b>Tarifas</b>	Un único tipo de tarifa sin diferenciación por clases y sin aplicación de descuentos por canal de contratación. Generalmente de cuantía inferior.		Diferenciación de clases de pasajeros y canales de contratación. Tarifas más complejas y generalmente superiores a las de LCC.
<b>Alianzas y códigos compartidos</b>	No		Si
<b>Venta</b>	Principalmente vía web y venta directa		Vía web, venta directa y agencias
<b>Aeronaves</b>	Modelo único y extensiva utilización		Varios modelos. Utilización inferior a las de las LCC

Fuente: E/CNMC/0002/14

Por su parte, los Gráfico 1 muestra el auge y desarrollo de las aerolíneas de bajo coste en el mercado mundial. De acuerdo con datos de la Comisión Europea, la evolución de este tipo de aerolíneas ha dado lugar al paso de una presencia en el mercado europeo de 1'5% en 1992 a prácticamente igualar la proporción de las compañías tradicionales en tan solo 20 años, presentando en todo caso porcentajes superiores a los del resto del mundo.



Fuente: "Visión 2050". IATA Febrero 2011.

La irrupción y auge de este nuevo modelo de aerolíneas ha tenido sus implicaciones sobre el desarrollo y la gestión de los aeropuertos. Éstos se han adaptado a las nuevas demandas de las mismas, a través, por ejemplo, de la eliminación de la obligatoriedad en la utilización de determinados servicios no indispensables (por ejemplo uso de *fingers* o autobuses) e incluso en último extremo dando lugar a una especialización de aeropuertos o terminales de aeropuertos para este tipo de compañías.

Adicionalmente al efecto de aparición y crecimiento de este tipo de compañías, también se observa en el mercado que las mismas presentan una mayor propensión al cambio de localización en los aeropuertos, derivada fundamentalmente de su mayor flexibilidad.

Tradicionalmente, la mayor parte de las aerolíneas establecían su base operativa en un aeropuerto concreto, convirtiéndose en cierto modo en aceptantes de las condiciones a ellas impuestas y con la presencia de un escaso margen de maniobra en sus actuaciones.

La dinámica actual ha dado lugar a una mayor facilidad de movimiento de dichas bases de las compañías aéreas. Este cambio de localización puede venir derivado de un abandono de la base de operaciones de la aerolínea, la quiebra de la compañía, o ser, fruto simplemente de una orientación premeditada de la misma hacia nuevas rutas origen-destino (o una reducción en las frecuencias de sus rutas). Como consecuencia, el ratio de creación y cierre de nuevas rutas entre pares de ciudades también ha experimentado un considerable aumento en los últimos años<sup>4</sup>. Además, estos efectos adquieren una mayor relevancia para los aeropuertos habida cuenta del tiempo que tarda el mismo en recuperarse del abandono de una compañía aérea.

No obstante, la modificación de estas rutas o, en su visión más extrema, la desaparición de una aerolínea en un determinado aeropuerto, también llevará asociados ciertos costes de cambio para éstas como los derivados de la recolocación de su personal, la ruptura de posibles contratos de largo plazo, el coste económico de las inversiones irre recuperables o una posible pérdida de las economías de escala en el caso de que el cambio suponga la división de las operaciones de la aerolínea en más de un aeropuerto.

Estas nuevas dinámicas suponen una limitación a la política comercial de calidad o precios que pueden desarrollar los gestores aeroportuarios, estando sujetos a la amenaza de posibles pérdidas de clientes (bien a través de la reducción de vuelos de las aerolíneas, bien por la desaparición por completo en el aeropuerto de la aerolínea) en el caso de que éstas no se adapten a las necesidades de las compañías.

---

<sup>4</sup> De acuerdo con “*Economics Report 2012*” de ACI, las aperturas y cierres de rutas de las aerolíneas europeas supusieron entre el 15 y el 20% del mercado.

La Tabla 2 constata la importancia de esta “amenaza” materializada en la reestructuración en las operaciones en los aeropuertos europeos de diferentes aerolíneas desde el año 2000 hasta la actualidad.

**Tabla 2. Reestructuración de hubs en Europa desde el año 2000**

<b>Aeropuerto</b>	<b>Aerolínea</b>	<b>Año</b>	<b>Motivo</b>
London Gatwick	British Airways	2000	Reducción de tamaño/reestructuración
Basilea	Swissair	2001	Quiebra
Bruselas	Sabena	2001	Quiebra y reapertura
Ginebra	Swissair	2001	Quiebra
Zurich	Swissair	2001	Quiebra y reapertura
Niza	Air Litoral	2001	Quiebra
Birmingham	British Airways	2003	Reducción de tamaño/reestructuración
Clemont-Ferrand	Air France	2004	Reducción de tamaño/reestructuración
Glasgow	British Airways	2006	Reducción de tamaño/reestructuración
Barcelona	iberia	2007	Reducción de tamaño/reestructuración
Milán Malpensa	Alitalia	2008	Reducción de tamaño/reestructuración
Atenas	Olympic Airlines	2009	Quiebra y reapertura
Manchester	British Midland	2009	Reducción de tamaño
Barcelona	Spanair	2012	Quiebra
Budapest	Malev	2012	Quiebra

Fuente: “Airport Competition”. Copenhagen Economics 2012.

Por tanto, nos encontramos ante un escenario diferente de relaciones entre aeropuertos y aerolíneas, en la medida en que diferentes aerolíneas y de forma destacada las de bajo coste, con modelos de gestión más flexible y fácilmente adaptable a otras bases, han ido ganando cuota de mercado.

Todo ello conduce en última instancia hacia una disminución del poder de mercado de los aeropuertos que aumentan su sensibilidad hacia los requerimientos de las aerolíneas en términos de desarrollo de sus políticas comerciales.

Para concluir podemos resumir que los elementos claves del modelo de las líneas aéreas de bajo coste en Europa apuestan, en general, por los siguientes componentes de la estrategia:

**- Tráfico aéreo de punto a punto, en lugar de vuelos de conexión en red y traslados a aeropuertos giratorios (hubs).**

Se rechazan conexiones de enlace con la misma aerolínea. De esta forma, se evitan los elevados costes resultantes de la complejidad del proceso de optimización de los trayectos de conexión en el sistema hub & spoke<sup>5</sup>.

Ofrecen solamente vuelos directos de corta y media distancia y con mucha frecuencia, que permiten una utilización máxima de la flota, renunciando a los trayectos intercontinentales. Rutas solamente con una alta participación del mercado y clase económica.

**- Vuelos sobre todo a aeropuertos secundarios o a antiguas bases militares.**

Las ventajas de la operación de estos tipos de aeropuertos pueden ser los bajos impuestos aeroportuarios o incluso ayudas financieras de instituciones públicas, como incentivos por traer visitantes que generan ingresos en la zona. Como consecuencia, las líneas aéreas de bajo coste se presentan como locomotoras económicas de ciertas regiones en Europa.

También, estos aeropuertos presentan un tamaño pequeño, con cortas vías de operación aeroportuaria y con menos tráfico aéreo, lo que ayuda a las líneas aéreas a evitar retrasos y reducir el tiempo de embarque y desembarque a un tiempo máximo de 25-30 minutos.

---

4 Centro y radios. El sistema de conexión de tráfico "hub & spoke" traducimos como de aporte y dispersión. Ejemplo: un aeropuerto C sirve como el *centro* (hub) de un sistema para llevar los pasajeros del aeropuerto A al aeropuerto B, en vez de llevarlos directamente de A a B, hay que realizarlo pasando por el centro C.

La operación en aeropuertos pequeños requiere también un inferior número de empleados para realizar las actividades necesarias para el funcionamiento de los vuelos.

**- En la mayoría de las aerolíneas de bajo coste no hay reservas de asiento.** Aunque en algunas se puede obtener una reserva de asiento contra pago adicional al efectuar la compra del vuelo, o en otras por ejemplo, conseguir la entrada preferente, es decir, poder subir a bordo del avión antes que los otros pasajeros.

**- No ofrecen ningún programa de puntos de pasajeros frecuentes ni salas de espera (lounges) para los pasajeros.**

Eliminan los periódicos y el servicio gratuito de catering a bordo. En lugar de ofrecer varias selecciones de menús calientes durante el vuelo, incluidos en el precio de billete, existe la posibilidad de compra de bebidas y otros productos alimenticios manufacturados contra pago extra.

**- La flota está estandarizada para obtener los efectos de sinergia en el mantenimiento y en la formación del personal.**

Las líneas aéreas de bajo coste apuestan por una flota moderna, lo que tiene ventajas tanto de seguridad como de reducción de costes por el consumo inferior de combustible y ahorros en los trabajos de mantenimiento.

Debido a que ofertan solamente vuelos de punto a punto sin enlace de conexión, se utilizan en general aparatos más pequeños.

En los aviones se encuentra un mayor número de asientos, gracias a la reducción del espacio entre ellos.

**- Reducción del tiempo de escala en tierra.**

A través del aumento del número de rotaciones de los vuelos, consiguen el mejor aprovechamiento de la flota, y así optimizan la utilización de la misma.

**- Plan de precios simple, las tarifas van incrementando a medida que se acerca la fecha de salida, los mejores precios se consiguen con mucha antelación.**

El uso de las nuevas tecnologías está muy intensificado, logrando mejores efectos de rendimiento de la venta.

**- Venta de billetes sobre todo vía Internet para evitar los gastos de comisiones de los intermediarios, agencias de viaje u otros sistemas de reservas.**

También se puede efectuar la compra del billete a través de llamada telefónica, no obstante, la cuota de reservas de esta forma es poco significativa, aproximadamente del 10 % y, además, produce un gasto adicional para el pasajero.

No emiten billetes en papel, los pasajeros reciben únicamente un número de reserva.

**- Los costes del personal se pretenden reducir a través de salarios inferiores a los de las líneas tradicionales por un lado, y por el otro, mediante una plena flexibilidad de empleo del personal en las tareas del proceso de operación relacionadas con el vuelo.**

Los gastos de las pernoctaciones del personal se pretenden evitar a través del estacionamiento de la flota en las bases, es decir el personal vuelve al origen del primer vuelo.

En estos casos no se trata solamente de una ventaja en el sentido de los gastos, sino también de la satisfacción de la demanda del personal de volver cada día a su propia casa y no pasar las noches en los hoteles.

**- Ingresos adicionales a la venta de los billetes de vuelo.**

A través de la comercialización de alojamientos en hoteles, alquileres de coches, seguros de viaje, billetes de transporte públicos y otros más. Éstos incrementan significativamente la cifra de facturación de la compañía aérea.

## 2. OBJETIVOS

El objetivo de este estudio es determinar si existe algún tipo de relación estadística, y analizar si ésta es modelizable linealmente, entre la variable “Staff (Número de empleados)” y las siguientes variables:

- ❖ Passenger: Pasajeros (número medio de kilómetros que cada pasajero ha sido transportado).
- ❖ Freight: Vuelos (toneladas-kilómetros volados).
- ❖ Fuel: Combustible (millones de galones).
- ❖ Other expenses: Otro tipo de gastos (en millones de dólares, excluyendo los gastos de mano de obra y los de combustible).
- ❖ Take off: Despegues (número total de despegues).

Por último, crearemos un modelo de carácter lineal que nos permita explicar o predecir el número de empleados de estas aerolíneas a través del resto de variables que hemos definido en el párrafo anterior.

## 3. METODOLOGÍA

Primero, utilizaremos gráficos de dispersión, el cálculo de correlaciones (simples y parciales) y el test de la t de Student para determinar la existencia o no de relaciones lineales entre las distintas variables.

Con objeto de establecer el mejor modelo lineal (regresión lineal múltiple), utilizaremos el estimador máximo verosímil de los parámetros del modelo, junto con el método Stepwise ligado al criterio AIK (Akaike Information Criteria).

Y por último, para analizar la bondad del ajuste haremos uso del error estándar residual, la tabla de ANOVA (Análisis de la Varianza) y el coeficiente de determinación. Junto a estos, analizaremos las hipótesis del modelo de regresión lineal múltiple (linealidad, normalidad, homocedasticidad, media cero e incorrelación) mediante las siguientes herramientas de análisis:

- ✓ El test de Shapiro-Wilks,
- ✓ El test de Breusch-Pagan,
- ✓ El test de Durbin-Watson,
- ✓ El histograma de los residuos,
- ✓ El gráfico qq-plot
- ✓ Otros gráficos habituales.

Para todos los análisis que llevemos a cabo, asumiremos un nivel de significación no superior al 5%.

El manual de referencia utilizado para la elaboración de este análisis estadístico es el manual Modelos Lineales Aplicados en R (Aparicio, Martínez Mayoral, & Morales, Modelos Lineales Aplicados en R), así como los video-tutoriales que aparecen como herramientas de apoyo de la asignatura de Econometría del profesor D. Juan Aparicio.

El programa utilizado en este estudio estadístico es R ([www.r-project.org](http://www.r-project.org)). R es un software libre (GNU), cuyo uso se ha generalizado en los últimos años entre estadísticos y usuarios de estas técnicas en todo el mundo.

### **3.1 Muestra de datos**

Hemos elaborado una base de datos del año 2009 con las variables mencionadas anteriormente referida a las aerolíneas del mundo (Ver tabla 3).

La fuente de información es la IATA (Asociación Internacional de Transporte Aéreo).

<b>airline</b>	Passenger	Freight	Staff	Fuel	Other_expenses	take_off
All Nippon Airways	35.261	614	12.222	860	2.008	6.074
Cathay Pacific (Hong-Kong)	23.388	1.580	12.214	456	1.492	4.174
Garuda (Indonesia)	14.074	539	10.428	304	3.171	3.305
JAL (Japan)	57.290	3.781	21.430	1.351	2.536	17.932
Malaysia Airlines	12.891	599	15.156	279	1.246	2.258
Quantas (Australia)	28.991	1.330	17.997	393	1.474	4.784
Saudia (Saudi Arabia)	18.969	760	24.078	235	806	6.819
SIA (Singapore)	32.404	1.902	10.864	523	1.512	4.479
AUA (Austria)	2.943	65	4.067	62	241	587
British Airways	67.364	2.618	51.802	1.294	4.276	12.161
Finnair (Finland)	9.925	157	8.630	185	303	1.482
Iberia (Spain)	23.312	845	30.140	499	1.238	3.771
Lufthansa (Germany)	50.989	5.346	45.514	1.078	3.314	9.004
SAS (Scandinavia)	20.799	619	22.180	377	1.234	3.119
Swissair	20.092	1.375	19.985	392	964	2.929
TAP Air Portugal	8.961	234	10.520	121	831	1.117
Air Canada	27.676	998	22.766	626	1.197	4.829
America West	18.378	169	11.914	309	611	2.124
American	133.796	1.838	80.627	2.381	5.149	18.624
Canadian	24.372	625	16.613	513	1.051	3.358
Continental	69.050	1.090	35.661	1.285	2.835	9.960
Delta	96.540	1.300	61.675	1.997	3.972	14.063
Eastern	29.050	245	21.350	580	1.498	4.459
Northwest	85.744	2.513	42.989	1.762	3.678	13.698
Pan American	54.054	1.382	28.638	991	2.193	7.131
TWA	62.345	1.119	35.783	1.118	2.389	8.704
United	131.905	2.326	73.902	2.246	5.678	18.204
Usair	59.001	392	53.557	1.252	3.030	8.952

Tabla 3. Base de Datos con las variables objeto de estudio.

## 4. ESTUDIO ECONÓMTRICO

### 4.1 ESTUDIO ECONÓMTRICO DEL MODELO COMPLETO

En nuestro estudio la variable respuesta es el número de NÚMERO DE EMPLEADOS (en adelante Staff) siendo las variables explicativas o covariables cinco indicadores económicos relacionados con la actividad de las aerolíneas. La variable respuesta la asumimos aleatoria, mientras que las variables explicativas carecen de aleatoriedad y sólo informan de lo que ocurre en los valores observados.

Con este estudio vamos a tratar de describir la relación entre la variable respuesta y las variables explicativas a través de un hiperplano, lineal en las variables explicativas, lo más próximo posible a los valores observados de la respuesta. No debemos olvidar, que la idea básica de la regresión múltiple es explicar lo máximo 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.

El modelo formulado para nuestro análisis es:

$$\text{Staff} = \beta_0 + \beta_1 * \text{Passenger} + \beta_2 * \text{Freight} + \beta_3 * \text{Fuel} + \beta_4 * \text{Other\_expenses} + \beta_5 * \text{Take\_off}$$

### **Definición de los parámetros:**

**$\beta_0$** : término de interceptación o intercepto.

**$\beta_1, \beta_2, \dots, \beta_5$**  son los coeficientes de regresión, son las pendientes de sus respectivas variables, por tanto el valor de  $\beta_1$  me indica que ceteris paribus, si aumento en una unidad la variable *Pasajeros* implica que **NÚMERO DE EMPLEADOS** aumentará en el valor de  $\beta_1$ . El valor del coeficiente  $\beta_1$  cuantifica el peso que tiene la variable *Pasajeros* explicando la respuesta media de **NÚMERO DE EMPLEADOS**. 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**

Lo primero que debemos hacer cuando queremos modelizar un banco de datos es llevar a cabo la inspección gráfica de los mismos, con el fin de determinar cómo son las relaciones entre las variables y si las hay. Así, en primer lugar analizaremos el gráfico de dispersión. Este, nos ofrecerá una primera aproximación a las relaciones existentes entre las variables.

En esta fase también realizaremos un análisis numérico de las correlaciones simples y parciales con objeto de tener una visión más objetiva sobre si la asociación se puede catalogar como lineal o no y nos permitirá cuantificar el grado de asociación lineal entre dichas variables.

El coeficiente de correlación parcial nos cuantificará el grado de relación lineal existente entre la variable respuesta y cada una de las variables explicativas, considerando siempre la información de las restantes variables a la hora de calcular el coeficiente.

Entrando ya en el análisis, la Figura 1 nos muestra todos los gráficos de dispersión posibles. Así, se deduce que al menos gráficamente existe una relación lineal “positiva” (creciente) entre el Staff y el resto de variables, destacando la variable Passenger (pasajeros) con mayor relación lineal aunque debemos puntualizar que en el caso de la variable Freight (vuelos) esta relación ser bastante menor.

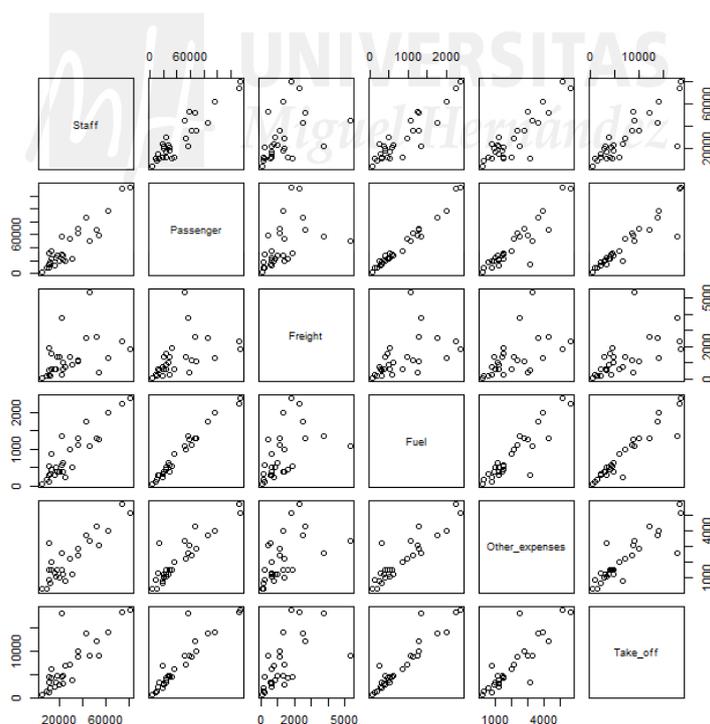


Figura 1 Relación entre las variables explicativas y la variable respuesta.

A continuación vamos a calcular los coeficientes de correlación y así poder asegurarnos numéricamente de que realmente existen relaciones lineales entre la variable respuesta y las variables explicativas.

VALORES DE LAS CORRELACIONES SIMPLES						
	Staff	Passenger	Freight	Fuel	Other_expenses	Take_off
Staff	1	0.9252213	0.4331636	0.9092500	0.8803991	0.8342433

Tabla 4 Valores correlación simple. Modelo completo

Analizando los resultados de la Tabla 4 comprobamos que efectivamente no existe una relación lineal entre la variable respuesta *Staff* (*empleados*) y la variable explicativa *Freight* (*vuelos*). Sin embargo, se confirman unas correlaciones simples elevadas entre *Passenger* y *Fuel*, siendo ambas el 0,925 y el 0,909 respectivamente. Esto nos indica que tienen una relación positiva, es decir, al aumentar dichas variables explicativas, la variable respuesta (*Staff*) también aumenta.

Con relación a las dos últimas variables, *Other\_expenses* y *Take\_off*, pese a que existe también una relación lineal positiva, su importancia dista bastante en comparación con las anteriores pues muestran una correlación simple de 0,880 y 0,834 respectivamente.

En cuanto a las correlaciones parciales, la Tabla 5 nos muestra los valores estimados. Observamos que la variable explicativa *Passenger* es la que tiene un mayor grado de relación lineal con la variable respuesta *Staff*, en segundo lugar la variable *Other\_expenser*, seguido de *Freight* y *Fuel*.

Cabe destacar que el valor obtenido por el *Freight* en las correlaciones simples es poco significativo, al igual que lo observado en la gráficas, esto se mantiene en los resultados determinados con el cálculo de las correlaciones parciales, pero en este caso su correlación parcial es superior a la de la variable *Fuel* que presentaba una correlación parcial superior y gráficamente mostraba una mayor relación lineal.

VALORES DE LAS CORRELACIONES PARCIALES						
	Staff	Passenger	Freight	Fuel	Other_expenses	Take_off
Staff	0	0.36613	0.03116	0.01646	0.23418	-0.18025

Tabla 5 Valores correlaciones parciales. Modelo completo

## FASE 2: AJUSTE DEL MODELO.

En esta fase vamos a ver los valores que estima el modelo para  $\beta_0$ ,  $\beta_1$ ,  $\beta_2$ ,  $\beta_3$ ,  $\beta_4$ ,  $\beta_5$ ,  $\beta_6$ , 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. Así, y en concreto, si se rechazase la hipótesis de coeficiente cero, sabríamos que esa variable es útil, es decir, contiene suficiente información para explicar la variable respuesta STAFF, en presencia del resto de covariables.

Los valores de los parámetros, por ejemplo  $\beta_1$ , 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 6 podemos ver 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:

$$\text{Staff} = 4206.5392 + 0.512 * \text{Passenger} + 0.2707 * \text{Freight} + 1.2374 * \text{Fuel} + 3.2180 * \text{Other\_expesnes} - 0.8662 * \text{Take\_off}$$

RESUMEN DEL MODELO						
Parametros	$\beta_0$	$\beta_1$	$\beta_2$	$\beta_3$	$\beta_4$	$\beta_5$
	4206.5392	0.5123	0.2707	1.2374	3.2180	-0.8662
p-valor	0.1576	0.0785	0.8851	0.9391	0.2707	0.3993
VALOR RESIDUAL ESTÁNDAR			8137			
Multiple R-squared			0.8689			
Ajusted R-squared			0.8391			
p-value			5.237e-09			

Tabla 6 Resumen del Modelo. Modelo completo.

La Tabla 7 se muestra los intervalos de confianza al 95% para cada uno de los coeficientes del modelo lineal.

INTERVALOS DE CONFIANZA		
	Extremo inferior	Extremo superior
INTERCEPTO	-1.756521e+03	10169.599665
Beta de Passenger	-6.340583e-02	1.088003
Beta de Freight	-3.569360e+00	4.110830
Beta de Fuel	-3.198769e+01	34.462425
Beta de Other_expe	-2.688918e+00	9.125.015
Beta de Take_off	-2.956143e+00	1.223753

Tabla 7 Intervalos de confianza, Modelo Completo

IC95%  $\beta_1$  = (-6.34, 1.08)

IC95%  $\beta_2$  = (-3.56, 4.11)

IC95%  $\beta_3$  = (-3.19, 3.46)

IC95%  $\beta_4$  = (-2.68, 9125.015)

IC95%  $\beta_5$  = (-2.95, 1.22)

A continuación, mediante un contraste de hipótesis estudiaremos si cada coeficiente puede ser o no igual a cero. En la Tabla 6 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 0.0785. 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. Para el resto de coeficientes obtendríamos:

p-valor  $\beta_2 = 0,8851 > 0,05$  No rechazamos  $H_0$ , asumiremos que el parámetro es nulo.

p-valor  $\beta_3 = 0,9391 > 0,05$  No rechazamos  $H_0$ , asumiremos que el parámetro es nulo.

p-valor  $\beta_4 = 0,2707 > 0,05$  No rechazamos  $H_0$ , asumiremos que el parámetro es nulo.

p-valor  $\beta_5 = 0,3993 > 0,05$  No rechazamos  $H_0$ , asumiremos que el parámetro es nulo.

Se observa que todos los p-valores son superiores a 0,05 por lo que no se puede rechazar  $H_0$ . En consecuencia, todas las beta 1, 2, 3, 4, 5 son cero lo que significa que la información contenida en cada una de estas variables, en presencia de las demás, no explica lo suficientemente bien la variable respuesta Staff y es probable que la información contenida en estas variables, se solapen entre ellas. En consecuencia no son variables importantes a la hora de explicar la variable respuesta Staff.

### FASE 3: BONDAD DEL AJUSTE

Para realizar el ajuste de 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 Número de empleados (Staff).

Hay tres medidas que nos van a permitir cuantificar la bondad del ajuste. Estas son las siguientes:

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

En nuestro caso el error estándar residual tiene un valor de 8137, y el coeficiente de variación de 28,38 % >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:

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

$H_1$ : 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  $5,237 \cdot 10^{-9}$  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 comprobar 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  $R^2 = 0,8689 > 65\%$ , por lo que el modelo es bastante bueno, y explica el 87 % de la variabilidad total de la variable Staff.

Por otra parte, el valor de coeficiente  $R^2$  ajustado = 0,8391, siendo prácticamente igual al coeficiente de  $R^2$  original. Por lo tanto, damos por válido el valor del coeficiente  $R^2$  original 0,8689, que como hemos indicado anteriormente es muy bueno.

#### **FASE 4: DIAGNÓSTICO DEL MODELO**

Después de ajustado 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:

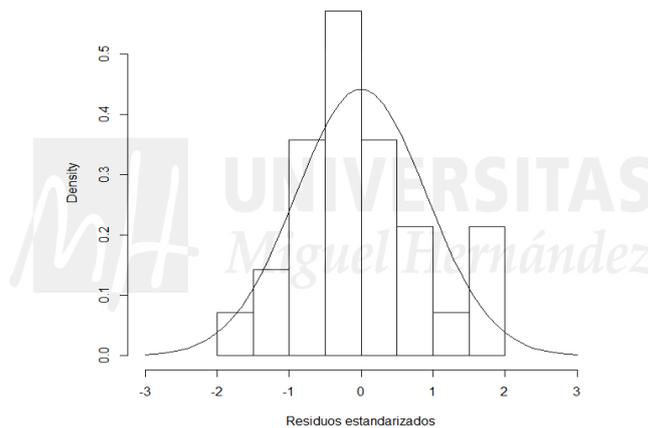
- Normalidad
- Varianza constante (Homocedasticidad)
- Incorrelación
- Media cero

Por otro lado, el análisis de los residuos nos permitirá estimar si existen o no deficiencias en la comprobación 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, realizando alguna transformación sobre las variables.

En esta fase del diagnóstico del modelo utilizaremos gráficos, así como 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.- 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.

**2.- Normalidad:** construiremos, en primer lugar, un histograma para comparar nuestro modelo con la campana de Gauss; a este respecto, 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.



*Figura 2 Histograma. Modelo completo*

En nuestro caso particular, Figura 2, el histograma no es exactamente una campana de Gauss, desviándose de la simetría y mostrando más peso en la parte central y al final de la cola de la derecha.

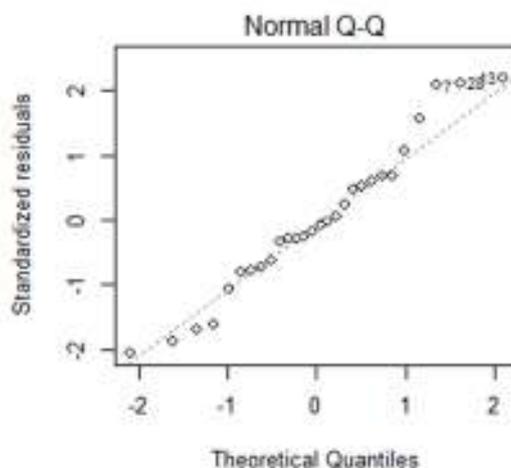


Figura 3 Gráfico QQPLOT. Modelo completo.

Por otra parte, en el gráfico qq-plot todos los puntos centrales están sobre la línea discontinua de referencia, mientras que en los puntos de los extremos se observa desviaciones, siendo superiores la de la derecha. Al igual que en el histograma, se observan desviaciones de la normalidad en ambas colas aunque de mayor relevancia en la cola de la derecha, por tanto, existen ciertas dudas sobre la normalidad.

A través del test de Shapiro-Wilks volveremos a contrastar la normalidad del modelo.

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

$H_1$ : lo contrario

El p-valor es de  $0,6925 > 0,05$ , por lo que no podemos rechazar la hipótesis de normalidad. En definitiva, esta hipótesis podemos asumir que se verifica.

**3.- Homocedasticidad (Hipótesis de Varianza Constante):** En primer lugar, elaboraremos un gráfico de residuos versus valores predichos de la variable respuesta Staff, y numéricamente resolveremos el contraste de hipótesis de Breusch-Pagan.

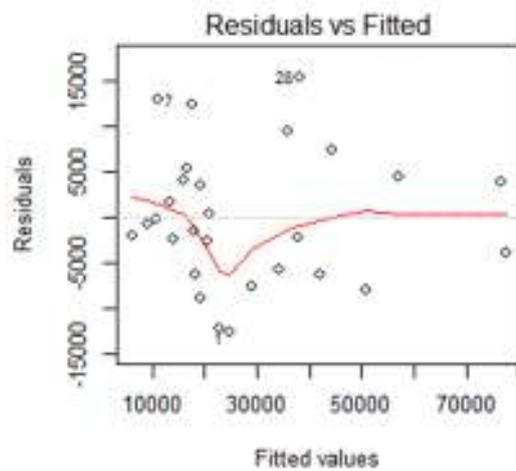


Figura 4 Residuos frente a valores predichos. Modelo completo.

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 4 vemos que la distribución de los puntos no tiene a lo largo de la gráfica la misma anchura (presenta forma de embudos que se va abriendo de derecha a izquierda), lo que nos indica que puede haber un problema respecto a la varianza constante. Para valores bajos de Staff (número de empleados) la variabilidad del error es muy elevada.

Desde el punto de vista gráfico, parece que no se cumple el criterio de homocedasticidad.

A continuación aplicaremos el test de Breusch y Pagan para resolver el contraste de hipótesis.

H0: Homocedasticidad

H1: Heterocedasticidad

En este caso, obtenemos un p-valor de  $0.6019 > 0,05$ . Por lo tanto, hay evidencias estadísticas para no rechazar  $H_0$ . Por lo que se cumple la hipótesis de homocedasticidad.

**3.- Incorrelación:** la incorrelación significa que las observaciones de la variable respuesta no están correlacionadas entre sí, es decir, los valores de unas variables no afectan a los de otras.

En primer lugar, para contrastar esta hipótesis, dibujaremos un gráfico del residuo  $i$  frente al anterior  $i-1$ .

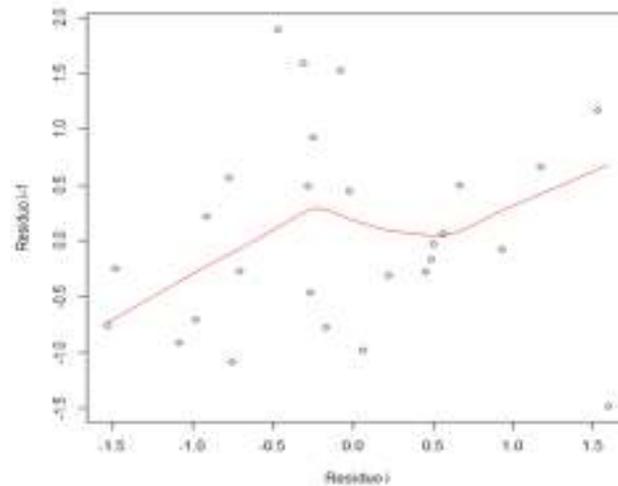


Figura 5 Gráfico de In correlación. Modelo completo.

En la figura 5 se observa una tendencia lineal creciente entre los errores. Esto, parece indicar que no se verifica la hipótesis de incorrelación. A continuación intentaremos comprobar dicha hipótesis de forma numérica a través de un p-valor obtenido a través del test de Durbin y Watson:

$H_0 : \rho = 0$  (incorrelación)

$H_1 : \rho \neq 0$

En este caso, obtenemos un p-valor= 0,187 > 0,05, por lo que no rechazamos  $H_0$ . Esto indica que el modelo es bueno para este criterio y asumiremos que sí se verifica la hipótesis de incorrelación.

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

PROCEDIMIENTO SECUENCIAL DE SELECCIÓN DE VARIABLES.

Mediante el criterio AIC y el método Stepwise vamos a determinar si en nuestro modelo original se puede reducir el 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 Original = 509.48	
VARIABLE	AIC
Fuel	507.49
Freight	507.51
Take_off	509.41
Other_expenses	509.06
Passenger	511.51

Tabla 8 Stepwise, paso 1

AIC Original = 507.49	
VARIABLE	AIC
Freight	505.52
Take_off	506.51
Other_expenses	507.11
Passenger	516.64

Tabla 9 Stepwise, paso 2

AIC Original = 505.52	
VARIABLE	AIC
Take_off	504.69
Other_expenses	505.38
Passenger	516.19

Tabla 10 Stepwise, paso 3

AIC Original = 504.69	
VARIABLE	AIC
Other_expenses	504.10
Passenger	516.59

Tabla 11 Stepwise, paso 4

AIC Original = 504.1	
VARIABLE	AIC
Passenger	556.37

Tabla 12 Stepwise, paso 5

El criterio AIC ya ha excluido todas aquellas variables que son menos relevantes de forma, que el nuevo modelo resultante nos queda siguiente forma:

$$\text{Staff} = \beta_0 + \beta_1 \text{ Passenger}$$

Con dicho criterio, la variable respuesta queda explicada únicamente a través de la variable Passenger. Por ello vamos a ver si mediante el método de significatividad estadística podemos obtener un modelo más sencillo (principio de parsimonia) distinto al obtenido con el criterio anterior.

#### 4.2-MÉTODO DE SIGNIFICATIVIDAD ESTADÍSTICA

MODELO COMPLETO						
Parametros	$\beta_0$	$\beta_1$	$\beta_2$	$\beta_3$	$\beta_4$	$\beta_5$
p-valor	0.1576	0.0785	0.8851	0.9391	0.2707	0.3993

Tabla 13 p-valores, modelo completo

P-VALOR ORIGINAL=0.9391					
Parametros	$\beta_0$	$\beta_1$	$\beta_2$	$\beta_4$	$\beta_5$
p-valor	0.14820	0.00274	0.88185	0.25329	0.36488

Tabla 14 p-valores, primer ajuste

P-VALOR ORIGINAL=0.88185				
Parametros	$\beta_0$	$\beta_1$	$\beta_4$	$\beta_5$
p-valor	0.1333	0.0011	0.2109	0.3208

Tabla 15 p-valores, segundo ajuste

<b>P-VALOR ORIGINAL=0.3208</b>			
Parametros	$\beta_0$	$\beta_1$	$\beta_4$
p-valor	0.172854	0.000484	0.266770

Tabla 15 p-valores, tercer ajuste

<b>P-VALOR ORIGINAL=0.26677</b>		
Parametros	$\beta_0$	$\beta_1$
p-valor	0.172854	0.000484

Tabla 16 p-valores, cuarto ajuste

El criterio significatividad estadística también ha excluido todas aquellas variables que son menos relevantes de forma, que el nuevo modelo resultante es igual al anterior.

$$\text{Staff} = \beta_0 + \beta_1 \text{ Passenger}$$

Como nada ha cambiado, a continuación vamos a realizar el análisis pormenorizado del nuevo modelo simplificado.

#### **4.3-ESTUDIO ECONÓMTRICO DEL MODELO SIMPLIFICADO**

Al igual que para el estudio completo de regresión múltiple vamos a analizar en las mismas cuatro fases el nuevo modelo simplificado:

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

##### **FASE 1: ANÁLISIS PRELIMINAR**

Como en el modelo completo vamos a llevar a cabo la inspección grafica de los nuevos datos que aparecen en la base de datos reducida a la variable explicativa Passenger. Así

podremos observar mediante el gráfico de dispersión que tipo de relación existe entre dicha variable y la variable respuesta.

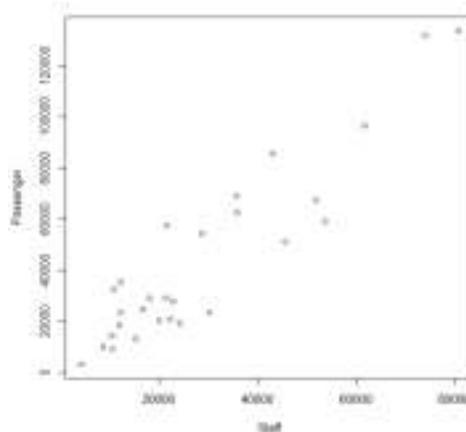


Figura 6 Relación entre la variable explicativa y la variable respuesta del Modelo Simplificado

La Figura 6 nos muestra el gráfico de dispersión del modelo reducido y se puede observar de nuevo una relación lineal entre Staff y Passenger, al igual que ocurría con el modelo completo, sigue existiendo una relación lineal creciente entre esta variable explicativa y la variable respuesta.

A continuación, vamos a realizar el análisis numérico para determinar las correlaciones simples y parciales con el fin de confirmar si realmente existe la asociación lineal que veíamos en el gráfico anterior.

VALOR DE LA CORRELACIÓN SIMPLE		
	Staff	Passenger
Staff	1	0.9252213

Tabla 17 Valor Correlación Simple, Modelo Simplificado.

Como podemos apreciar en los resultados de la Tabla 17, el valor obtenido es igual que en el modelo completo bastante cercano a 1, con lo cual, se vuelve a confirmar la relación lineal existente entre la variable Staff y la variable explicativa Passenger.

En la Tabla 18, podemos ver que en el modelo simplificado, la correlación parcial es de 0.925 muy superior al valor obtenido en el modelo completo que era de 0.366.

<b>VALOR DE LA CORRELACIÓN PARCIAL</b>		
	Staff	Passenger
Staff	1	0.92522

Tabla 18 Valor Correlación Parcial, Modelo Simplificado.

## FASE 2: AJUSTE DEL MODELO

En esta segunda fase veremos los valores que estima el modelo para  $\beta_0$  y  $\beta_1$ , 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 rechacemos la hipótesis de coeficiente cero determinaremos si la variable Passenger contiene suficiente información para explicar la variable respuesta Staff. Estimaremos el valor de los parámetros con el criterio de máxima verosimilitud.

<b>RESUMEN DEL MODELO</b>		
<b>Parametros</b>	$\beta_0$	$\beta_1$
	5.231	5.381
p-valor	0.0384	$1.9 \cdot 10^{-12}$
VALOR RESIDUAL ESTANDAR	7843	
Multiple R-squared	0.856	
Ajusted R-squared	0.8505	
p-value	1.9e-12	

Tabla 19 Resumen del Modelo Simplificado

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

$$\text{Staff} = 5,231 + 5,381 \cdot \text{Passenger}$$

Con este nuevo modelo observamos una relación directa entre Staff y Passenger.

En la Tabla 20 muestra el intervalo de confianza al 95%.

INTERVALOS DE CONFIANZA		
	Extremo inferior	Extremo superior
INTECEPTO	301.8669018	1.015981e+04
Beta de Passenger	0.4491361	6.270492e-01

Tabla 20 Intervalos de confianza. Modelo Simplificado

A continuación vamos a realizar el contraste de hipótesis para comprobar si el coeficiente contiene suficiente información para explicar la variable explicativa.

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

H0:  $\beta_1 = 0$

H1:  $\beta_1 \neq 0$

El p-valor de este par de contraste-coeficiente es igual a  $1,9 \cdot 10^{-12}$ , inferior a 0,05 por tanto rechazamos H0 y asumimos que el  $\beta_1$  es distinto de cero.

$\beta_1$  es útil por contener información que explica la variable Staff.

### FASE 3: BONDAD DEL AJUSTE

Al igual que en el 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 (Staff).

Para ello utilizaremos tres criterios:

**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 19 obtenemos el valor residual estándar del modelo, el cual tiene un valor de 7843. Para determinar si este valor es grande o pequeño realizamos el cálculo del coeficiente de variación  $cv = 100 \cdot (7843 / \text{media de Staff}) \%$  lo que da un resultado de  $cv = 27.35\% > 10\%$ . Por lo tanto el modelo NO es bueno en base a este criterio, aunque mejora en un 1% con respecto al modelo completo.

**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:

H0:  $\beta_1=0$

H1: lo contrario

El p-valor de nuestro modelo es  $1,9 \cdot 10^{-12}$  que como es menor que 0,05, rechazo H0 y asumimos que el modelo es bueno al igual que en el modelo completo.

**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.

En este modelo simplificado, como solo existe una covariable solo tendremos en cuenta el valor del coeficiente de R2.

En nuestro caso el coeficiente de R2 es 0,856 ( $\approx 100\%$ ), y por tanto la recta de regresión parece explicar la variabilidad total de los datos. Podemos decir que el modelo es bueno.

Anteriormente en el modelo completo tenía un valor del 86,8%, por lo que este modelo simplificado sigue explicando bien la variabilidad de la variable respuesta.

#### **FASE 4: DIAGNÓSTICO DEL MODELO**

Tras haber superado las fases anteriores, excepto el criterio del error estándar residual, 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. Si no se cumpliera alguna de las siguientes hipótesis (normalidad, homocedasticidad, varianza constante e incorrelación), tendríamos que rechazar el modelo simplificado y nos replantearíamos un modelo empleando alguna transformación sobre las variables.

Para estudiar 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 gráfico QQPLOT donde representamos los residuos ordenados versus los cuantiles correspondientes de una normal y el histograma como en el modelo anterior.

En el gráfico qqplot, Figura 7, casi de los puntos del centro están sobre la línea recta, pero en las colas se desvían lo que parece indicar que pueden existir problemas de normalidad.

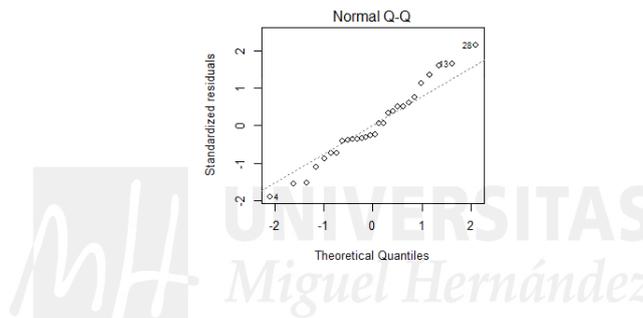


Figura 7 Gráfico QQPLOT. Modelo Simplificado

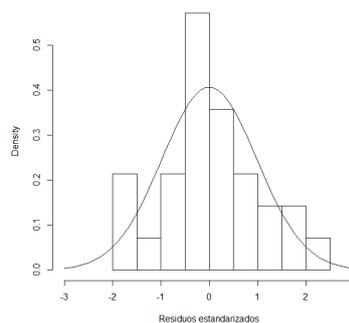


Figura 9 Histograma. Modelo Simplificado

En el histograma, Figura 8, vemos que la mayoría 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, al igual que en la figura anterior solo algunos valores centrales parecen cumplirla.

Ante esta situación de incertidumbre sobre la existencia de normalidad vamos a realizar una comprobación numérica mediante el contraste de hipótesis de Shapiro-Wilks:

H0:  $e \sim \text{Normal}$

H1: lo contrario

El valor obtenido es un p-valor de  $0.7853 > 0,05$  por lo tanto, asumimos en definitiva, 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 Staff y numéricamente resolveremos el contraste de hipótesis de Breush-Pagan.

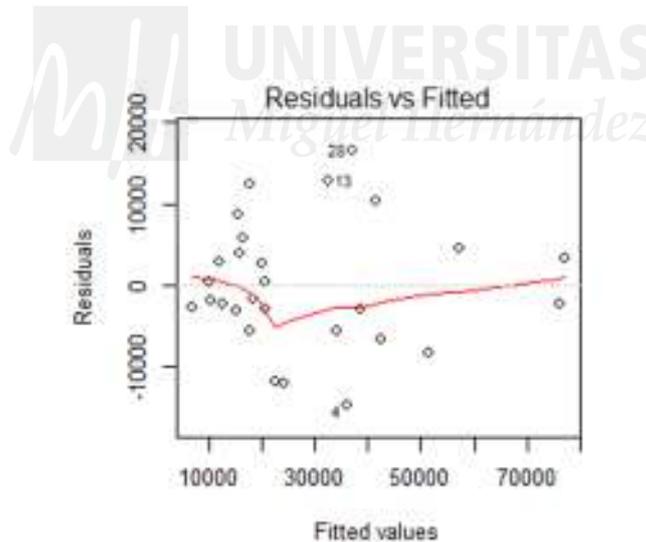


Figura 10 Residuos frente a valores predichos. Modelo Simplificado

En el gráfico de la Figura 10, vemos que la distribución de los puntos no tiene la misma anchura a lo largo de la gráfica, a la izquierda es mucho más ancho, estrechándose progresivamente desde el centro hacia 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.

Comprobaremos numéricamente a través del contraste de hipótesis de Breusch-Pagan si se cumple la Homocedasticidad.

H0: Homocedasticidad

H1: Heterocedasticidad

En este caso obtenemos un p-valor de  $0.5808 > 0,05$ , por lo que no podemos rechazar H0 lo que significa que el modelo cumple con la hipótesis de varianza constante, a pesar de que gráficamente parecía no cumplirse.

**3.- Incorrelación:** vamos a establecer si los valores de unas variables afectan a las de otras, para comprobarlo construiremos el gráfico del residuo  $i$  frente al anterior  $i-1$  y realizaremos el contraste de hipótesis de Durbin y Watson para hacer la comprobación numérica.

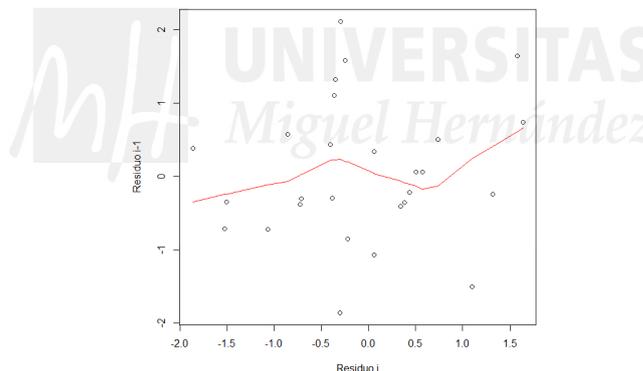


Figura 11 Gráfico de Incorrelación. Modelo Simplificado

En el gráfico de la Figura 11 vemos una tendencia lineal creciente (la línea roja superpuesta de la figura) entre los errores. Esto en principio puede indicar que no se cumple la hipótesis de incorrelación.

Para realizar la comprobación numérica de esto vamos a realizar el contraste de Durbin-Watson:

H0 :  $\rho = 0$  (se cumple el criterio de Incorrelación)

H1:  $\rho \neq 0$  (No se cumple el criterio de Incorrelación)

Obtenemos un p-valor =  $0.2196 > 0,05$  que nos 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 éstas mejoran, principalmente los gráficos del diagnóstico de los residuos.

Inicialmente, para determinar qué transformación es aceptable nos apoyaremos en un método de transformación que pertenece a la familia de transformaciones potenciales, como es el caso de las transformaciones Box-Cox. Este método utiliza como elemento determinante el resultado que arroja el valor lambda.

Cuando el valor de lambda es distinto de cero es recomendable que la transformación sea la variable respuesta elevada al valor lambda, y cuando lambda es 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.

En nuestro estudio, hemos obtenido mediante la transformación de Box-Cox un valor de lambda de 0,7. Dicho valor, próximo a 1 nos está indicando que el modelo original es bueno.

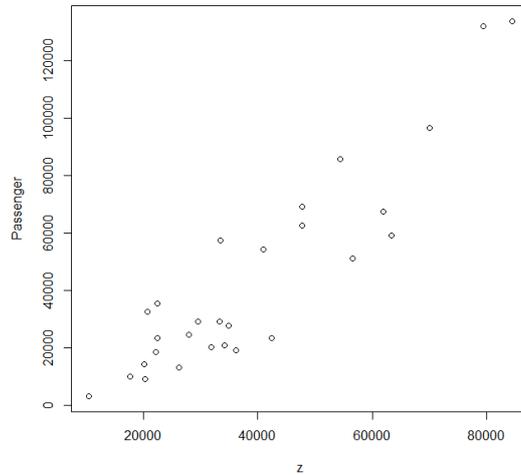


Figura 12 Relación entre la variable explicativa y z.

### FASE 1: ANÁLISIS PRELIMINAR

Del análisis gráfico de la Figura 12 podemos observar, al igual que en los anteriores modelos, que existe una fuerte relación lineal “positiva” (creciente) entre la nueva variable transformada Z y la covariable Passenger.

VALORES DE LAS CORRELACIONES SIMPLES		
	z	Passenger
<b>Staff</b>	1	0.9133071

Tabla 21 Valores de las Correlaciones Simples. Modelo Transformado

La tabla correlaciones simples (Tabla 21), nos vuelve a indicar al igual que en la tabla 4 y en la gráfica anterior que existe una correlación simple elevada entre la variable transformada y la variables Passenger.

VALORES DE LAS CORRELACIONES PARCIALES		
	z	Passenger
<b>Staff</b>	0	0.91331

Tabla 22 Valores de las Correlaciones Parciales. Modelo Transformado

En la Tabla 22 aparecen representados los valores de las correlaciones parciales, nuevamente la variable “passenger” vuelve a presentar un fuerte grado de relación lineal con un valor de 0,9133 igual al de la correlación simple.

## FASE 2: AJUSTE DEL MODELO

$$\text{Staff} = 5231 + 5.381 * \text{Passenger}$$

RESUMEN DEL MODELO		
	$\beta_0$	$\beta_1$
Parametros	$5.231 * 10^3$	$5.381 * 10^{-1}$
p-valor	0.0384	$1.9 * 10^{-12}$
Valor residual estándar	7843	
Multiple R-squared	0.856	
Ajusted R-squared	0.8505	
p-value	$1.9 * 10^{-12}$	

Tabla 23 Resumen del modelo. Modelo Transformado

Los intervalos de confianza son para cada uno de los coeficientes son:

INTERVALOS DE CONFIANZA		
	Extremo inferior	Extremo superior
INTERCEPTO	301.8669	$1.015981 * 10^4$
Beta de Passenger	0.449136	$6.270492 * 10^{-1}$

Tabla 24 Intervalos de confianza. Modelo Transformado

A continuación vamos a realizar un contraste de hipótesis y poder comprobar si el coeficiente contiene información lo suficientemente útil como para explicar la variable

explicativa. En la Tabla 23 están recogidos los p-valores asociados para cada par de contraste-coeficiente.

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

H0:  $\beta_1 = 0$

H1:  $\beta_1 \neq 0$

El p-valor es igual a  $1.9 \cdot 10^{-12}$ , inferior a 0,05 por tanto rechazamos H0 y asumimos que el parámetro no es cero. Esto nos está indicando que la variable Passenger es lo suficientemente útil para explicar Staff.

### FASE 3: BONDAD DEL AJUSTE

**1.- Error estándar residual:** De los datos contenidos en la Tabla 23, podemos ver que el valor residual estándar del modelo es 7843. Para determinar si este valor es grande o pequeño realizamos el cálculo del coeficiente de variación cuya fórmula es la siguiente:

$cv = 100 \cdot (7843 / \text{la media del número de empleados transformada}) \%$  .

Esto nos da un resultado de  $cv = 20,10$ . Al ser mayor del 10% nos indica que el modelo no es bueno, aunque el valor se ha reducido respecto del modelo simplificado que era 27,35. El nuevo modelo no mejora de manera significativa el modelo simplificado.

**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:

H0:  $\beta_1 = 0$

H1: lo contrario

El p-valor de nuestro modelo es  $1.9 \cdot 10^{-12}$  que como es menor que 0,05, rechazo  $H_0$  y asumimos que el modelo es bueno. La variable explicativa explica suficientemente bien la variable respuesta a través del modelo lineal propuesto.

**3.- Coeficiente de determinación:** nos indicará el valor en tanto por ciento de la variabilidad total de los datos. Esto es la proporción de la varianza que es explicada por la recta de regresión. Por ello, cuanto más cerca de 1 (100%) esté 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.

El coeficiente de  $R^2$  es 0.856 ( $\approx 100\%$ ), parece explicar la recta de regresión., por lo que podemos decir que el modelo es bueno.

Este modelo transformado sigue explicando bien la variabilidad de la variable respuesta.

#### FASE 4: DIAGNÓSTICO DEL MODELO

**1.- Normalidad:** Al igual que hicimos en el modelo anterior vamos a realizar un estudio gráfico viendo el gráfico QQPLOT donde representamos los residuos ordenados versus los cuantiles correspondientes de una normal y el histograma.

En el gráfico qqplot (Figura 13), la mayoría de los puntos del centro están sobre la línea recta, sin embargo en las colas se sigue manteniendo la desviación, principalmente en la cola de la derecha. Esto nos indica que puede haber problemas de normalidad.

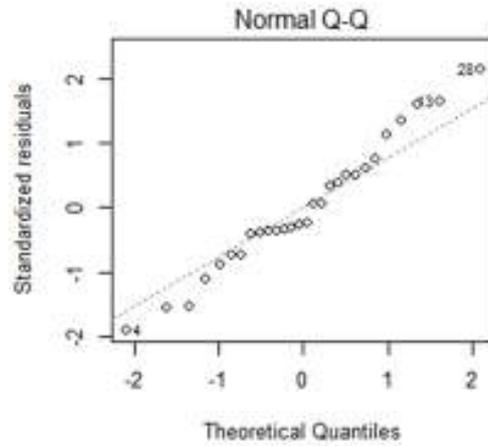


Figura 13 Gráfico QQ-PLOT. Modelo Transformado

Vemos que en el histograma (figura 14) falla la normalidad. Existe una clara desviación de la simetría, principalmente en la parte central y en la derecha. Este modelo transformado no ha mejorado respecto al modelo simplificado analizado anteriormente.

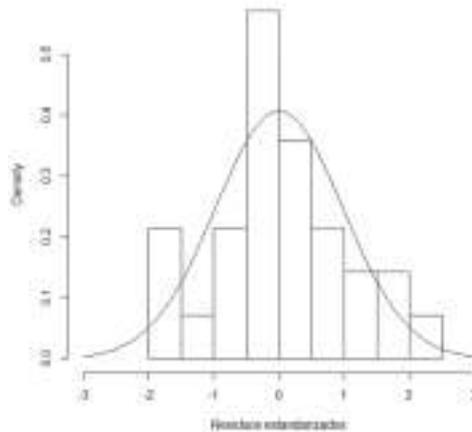


Figura 14 Residuos Estandarizados. Modelo transformado

Como este criterio grafico no es suficiente para determinar la normalidad, vamos a realizar una comprobación numérica mediante el contraste de hipótesis de Shapiro-Wilks.

H0:  $e \sim \text{Normal}$

H1: lo contrario

El valor obtenido es un p-valor de  $0.7853 > 0,05$  por lo tanto, asumimos en definitiva, que sí se verifica la hipótesis de normalidad, al igual que en el modelo anterior cuyo p-valor era de 0.6243.

**2.- Homocedasticidad: (hipótesis de varianza constante).** Al igual que en el criterio anterior, realizaremos un análisis sobre el gráfico de residuos versus valores predichos de la variable respuesta Staff (transformada) y numéricamente resolveremos el contraste de hipótesis de Breush-Pagan.

En el siguiente gráfico, vemos que la distribución de los puntos no tiene a lo largo de la gráfica la misma anchura, a la derecha es donde parece existir una anchura similar. 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. Poco mejora, el modelo transformado respecto del simplificado.

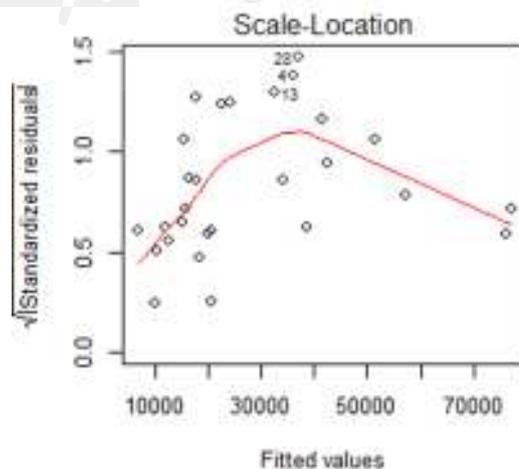


Figura 15 Residuos frente a valores predichos. Modelo Transformado

Corresponde ahora asegurar el cumplimiento de ésta hipótesis a través del contraste de hipótesis de Breusch-Pagan:

H0: Homocedasticidad

H1: Heterocedasticidad

Obtenemos un p-valor de  $0.5808 > 0,05$ , por lo que no podemos rechazar  $H_0$  lo que implica que asumimos que el modelo cumple con la hipótesis de varianza constante.

**3.- Incorrelación:** según este criterio vamos a determinar si los valores de unas variables afectan a las de otras.

Para su comprobación construiremos el gráfico del residuo  $i$  frente al anterior  $i-1$  y el contraste de hipótesis de Durbin y Watson.

En la Figura 16 vemos una tendencia lineal creciente (la línea roja superpuesta de la figura) entre los errores, por lo que parecería que la hipótesis de incorrelación NO se cumple.

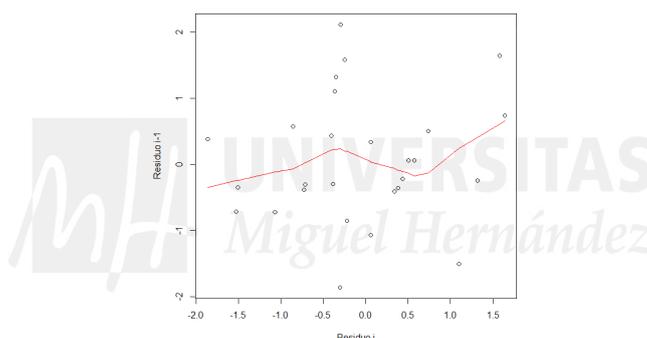


Figura 16 Gráfico de Incorrelación. Modelo transformado

Comprobaremos que resultados se obtienen con el test de Durbin-Watson.

$H_0 : \rho = 0$  (se cumple el criterio de Incorrelación)

$H_1: \rho \neq 0$  (No se cumple el criterio de Incorrelación)

Se obtiene un p-valor =  $0.2196 > 0,05$ . Esto nos indica que el modelo es bueno y asumimos que cumple el criterio de Incorrelación, manteniéndose el mismo p-valor que obtuvimos en el modelo simplificado.

**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 simplificado, el nuevo modelo conseguido mediante la transformación logarítmica de nuestra variable respuesta "Staff", cumple con las bondades del ajuste y se verifican las hipótesis asumidas por el modelo, mejorando ligeramente (en alguno de los criterios) los diagnósticos gráficos de verificación de las hipótesis.

## 6. CONCLUSIONES

En este estudio, aunque hemos analizado tres modelos distintos, estos están estrechamente relacionados, un primer análisis con el modelo lineal completo, y a continuación, un segundo modelo en el que hemos llevado a cabo un análisis simplificado, donde hemos prescindido de aquellas variables que no aportaban información relevante para explicar la variable respuesta; y ya por último un tercer modelo donde hemos transformado la variable respuesta "Staff".

Con ello se ha pretendido comprobar si existe algún tipo de relación lineal entre la variable respuesta y las variables explicativas descritas anteriormente.

No obstante, resaltar que como con el criterio AIC, el modelo reducido que se obtiene contenía una sola covariable, hemos decidido aplicar el método de significatividad estadística para ver si era posible obtener un modelo distinto a este. Sin embargo, como nada ha cambiado, hemos considerado oportuno realizar el análisis de modelo simplificado.

Para el análisis de los tres modelos hemos seguido el mismo procedimiento o criterios en cada uno de ellos. En primer lugar hemos realizado un análisis preliminar, a continuación hemos ajustado el modelo, comprobado la bondad de los ajustes y por último hemos verificado que se cumplen las hipótesis asumidas por el modelo.

En el análisis preliminar, en primer lugar, hemos analizado gráfica y numéricamente las relaciones existentes entre las variables explicativas y nuestra variable respuesta.

Los gráficos de dispersión nos han mostrado una relación lineal "positiva" (creciente) entre el Staff y el resto de variables, destacando la variable Passenger (pasajeros) aunque debemos puntualizar que en el caso de la variable Freight (vuelos) esta relación es menos importante (caso modelo completo).

Con el análisis de las correlaciones simples corroboramos lo observado gráficamente y efectivamente no existe una relación lineal entre la variable respuesta *Staff* (*empleados*) y la variable explicativa *Freight* (*vuelos*). Sin embargo, se confirman unas correlaciones simples elevadas de Passenger y Fuel. Esto nos indica que tienen una relación positiva, es decir, al aumentar dichas variables explicativas, la variable respuesta (*Staff*) también aumenta. Sin embargo, las dos últimas variables, *Other\_expenses* y *Take\_off*, pese a que existe también una relación lineal positiva, su importancia dista bastante en comparación con las anteriores.

En cuanto a las correlaciones parciales, observamos que la variable explicativa Passenger es la que tiene un mayor grado de relación lineal con la variable respuesta Staff, en segundo lugar la variable *Other\_expenses*, seguido de Fuel y Freight, pese a que esta última parecía no tener mucha importancia en el modelo.

En el modelo simplificado, excepto la variable explicativa Passenger, todas las demás son eliminadas por carecer de valor explicativo.

Una vez realizado el análisis preliminar hemos procedido a ajustar el modelo para estimar los valores de los seis parámetros en el modelo completo y de un parámetro en el modelo simplificado. También hemos calculado sus intervalos de confianza al 95%.

El modelo reducido, queda definido de la siguiente manera: **Staff** = 5,231+ 5,381\*Passenger. Esto significa que si el número medio de kilómetros que cada pasajero ha sido transportado aumenta en 1000 unidades (kilómetros) el personal de plantilla se ve incrementado en 5381 unidades.

En el primer modelo, todos los p-valores son superiores a 0.05 por lo que en ningún caso se puede rechazar H0. En consecuencia, toda la información contenida en las variables explicativas no explica lo suficientemente bien la variable respuesta Staff.

En el modelo simplificado ocurre todo lo contrario ya que la variable contiene información suficientemente útil para explicar la variable respuesta.

Posteriormente hemos realizado la bondad del ajuste, y así poder verificar si nuestro modelo es bueno para explicar la variable respuesta. Esto se comprueba utilizando los criterios del Error Estándar Residual, la Tabla de Anova y el Coeficiente de Determinación.

En ambos modelos se obtiene lo mismo, esto es, que el criterio del Error Estándar Residual es desfavorable mientras que los restantes dos criterios han sido favorables, indicándonos que los modelos son buenos.

Por último, hemos llevado a cabo un diagnóstico de los modelos para comprobar si éstos cumplen las hipótesis básicas de regresión lineal para los errores. Desde el punto de vista gráfico hemos utilizado las siguientes gráficas:

- Histograma y gráfico qqplot en la normalidad,
- Residuos frente a valores predichos en la homocedasticidad
- Gráfico de incorrelación para las incorrelaciones

Desde el punto de vista numérico hemos usado como contrastes de hipótesis:

- Shapiro-Wilks en la normalidad,
- Breusch-Pagan en la homocedasticidad
- Durbin-Watson para la incorrelación).

La hipótesis de media cero se cumple siempre por usar el método de máxima verosimilitud.

En los dos modelos, los gráficos 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.

Por último, decir que como el coeficiente asociado a la variable Passenger tiene signo positivo, podemos afirmar que si aumenta el número medio de kilómetros que cada pasajero ha sido transportado, entonces necesariamente aumenta el número de empleados.

## BIBLIOGRAFIA

- Aparicio, J. (s.f.). *ECONOMETRÍA*. Recuperado el 05 de mayo de 2015, de <http://umh3067.edu.umh.es/material/videos/>
- Aparicio, J. M. (s.f.). *Modelos Lineales Aplicados en R*. (C. I. Hernández, Ed.) Dto. Estadística, Matemáticas e Informática.
- CNMC. (2014). *El sector aeroportuario en España: Situación actual y recomendaciones de liberalización*. Recuperado el 06 de mayo de 2015, de [http://www.cnmc.es/Portals/0/Notas%20de%20prensa/20140703\\_Estudio\\_aeroportuario\\_%20integrado.pdf](http://www.cnmc.es/Portals/0/Notas%20de%20prensa/20140703_Estudio_aeroportuario_%20integrado.pdf)
- Gross, S. y. (2004). Low Cost Airlines in Europa. *Erich Schmidt Verlag GmbH & Co., Berlin.*, 3-4.
- Meissner, M. (2008). *Las líneas aéreas de bajo coste: Fundamentos teóricos y estudio empírico sobre su impacto en el transporte aéreo y en la estructura del sector turístico en Europa*. Recuperado el 05 de mayo de 2015, de <http://dialnet.unirioja.es/servlet/tesis?codigo=21955>

