



**TESIS DE GRADO
EN INGENIERÍA INDUSTRIAL**

**HERRAMIENTAS PARA LA MEJORA DE LA
GESTIÓN COMERCIAL DEL TURISMO EMISIVO**

Autor: Hernán José Piqué

Director de Tesis: Ing. Roberto Mariano García

2007

Dedico este proyecto a mis familiares, que me inculcaron desde niño la importancia de emprender una carrera universitaria. Su ayuda y consejos fueron imprescindibles en el desarrollo de la misma, por lo cual les estaré eternamente agradecido.

RESUMEN

Este proyecto propone la elaboración de modelos de apoyo que puedan utilizar las agencias de viajes u otros en el rubro turismo. Se trata de, entre otras cosas, determinar si una cierta persona es cliente potencial de algún tipo especial de viaje de placer o, de ser un cliente existente, si estaría interesado en algún paquete turístico en particular. Para tal fin, la regresión logística es la técnica estadística que mejor se adapta al tipo de datos relevados.

A su vez, se trata de demostrar que la aplicación de los modelos obtenidos mejoran los resultados económicos de las áreas comerciales y, por ende, de la industria del turismo en general.

SUMMARY

This project proposes the elaboration of support models that could be used by travel agencies or others in the tourism field. It is tried, among other things, to determine if a certain person is a potential client of some special type of pleasure trip or, of being an existing client, if it would be interested in some tourist package in individual. For such aim, the logistic regression is the statistical technique that better adapts to the investigated data type.

As well, it is shown that the application of the obtained models improves the economic results of the commercial areas and, therefore, of the industry of tourism in general.

TABLA DE CONTENIDOS

SECCIÓN I.....	1
CAPÍTULO 1 - INTRODUCCIÓN.....	1
CAPÍTULO 2 - OBJETIVOS, MARCO TEÓRICO Y MODELO CONCEPTUAL.....	9
CAPÍTULO 3 - MUESTRA.....	11
CAPÍTULO 4 - PROCEDIMIENTOS ESTADÍSTICOS.....	17
CAPÍTULO 5 - PASOS PARA LA OBTENCIÓN DE UN MODELO DE REGRESIÓN LOGÍSTICA BINARIA.....	19
5.1 Estimación de los Parámetros por Máxima Verosimilitud.....	19
5.2 Validación del Modelo.....	22
CAPÍTULO 6 - PASOS PARA LA OBTENCIÓN DE UN MODELO DE REGRESIÓN LOGÍSTICA MULTINOMIAL.....	25
6.1 Indagaciones Exploratorias previas.....	25
6.1.1 Análisis de Correlaciones Bivariadas.....	25
6.1.2 Análisis de Contrastes de Razón de Verosimilitud.....	26
6.1.3 Evaluar Ausencia de Colinealidad entre las Variables Independientes.....	27
6.2 Planteo de Hipótesis.....	28
6.3 Resultados del modelo.....	28
6.3.1 Estimación de Los Coeficientes y Evaluación del Modelo Obtenido.....	28
6.3.2 Contraste de Hipótesis.....	34
SECCIÓN II.....	39
CAPÍTULO 7 - MODELOS OBTENIDOS A PARTIR DE LA INVESTIGACIÓN..	41
7.1 Modelos Generales referentes a Viajes Turísticos.....	41
7.1.1 Modelo Referente a la Elección de Tipo de Hospedaje.....	41
7.1.1.1 Planteo de Hipótesis.....	41
7.1.1.1 Resultados del Modelo.....	42
7.1.1.2 Contraste de Hipótesis.....	44
7.1.1.3 Resumen y Conclusiones del Modelo.....	45
7.1.2 Modelo Referente a la Elección de Distintos Destinos.....	45
7.1.2.1 Planteo de Hipótesis.....	46
7.1.2.2 Resultados del Modelo.....	47
7.1.2.3 Contraste de Hipótesis.....	49
7.1.2.4 Resumen y Conclusiones del Modelo.....	50
7.1.3 Modelo Referente a la Elección del Destino Europa.....	51
7.1.3.1 Planteo de Hipótesis.....	52
7.1.3.2 Resultados del Modelo.....	52
7.1.3.3 Contraste de Hipótesis.....	54
7.1.3.4 Resumen y Conclusiones del modelo.....	55
7.1.4 Modelo Referente a la Elección de la Modalidad de Organización.....	56
7.1.4.1 Planteo de Hipótesis.....	57
7.1.4.2 Resultados del Modelo.....	57
7.1.4.3 Contraste de Hipótesis.....	59
7.1.4.4 Resumen y Conclusiones del Modelo.....	61

7.2	Modelos para Jóvenes referentes a Viajes Turísticos	62
7.2.1	Modelo Referente a la Elección de Tipo de Hospedaje	62
7.2.1.1	Planteo de Hipótesis	62
7.2.1.2	Resultados del Modelo	63
7.2.1.3	Contraste de Hipótesis	65
7.2.1.4	Resumen y Conclusiones del Modelo	65
7.2.2	Modelo Referente a la Elección de Distintos Destinos	66
7.2.2.1	Planteo de Hipótesis	67
7.2.2.2	Resultados del Modelo	68
7.2.2.3	Contraste de Hipótesis	69
7.2.2.4	Resumen y Conclusiones del Modelo	70
7.2.3	Modelo Referente a la Elección del Destino Europa	71
7.2.3.1	Planteo de Hipótesis	72
7.2.3.2	Resultados del Modelo	72
7.2.3.3	Contraste de Hipótesis	74
7.2.3.4	Resumen y Conclusiones del modelo	75
7.2.4	Modelo Referente a la Elección de la Modalidad de Organización	76
7.2.4.1	Planteo de Hipótesis	77
7.2.4.2	Resultados del Modelo	77
7.2.4.3	Contraste de Hipótesis	79
7.2.4.4	Resumen y Conclusiones del Modelo	81

SECCIÓN III..... 83

CAPÍTULO 8 – APLICACIONES PRÁCTICAS DE LOS MODELOS DE REGRESIÓN OBTENIDOS..... 85

8.1	Comparación de Alternativas de envío de Publicidad de Viajes con Hoteles de Lujo para Jóvenes.....	85
8.1.1	Alternativa 1	86
8.1.2	Alternativa 2	87
8.1.3	Alternativa 3	88
8.1.4	Comparación de Resultados	89
8.2	Comparación de Alternativas de envío de Publicidad de Viajes con Distintos Destinos para Jóvenes.....	92
8.2.1	Alternativa 1	93
8.2.2	Alternativa 2	94
8.2.3	Alternativa 3	94
8.2.4	Comparación de resultados.....	95
8.3	Comparación de Alternativas de Publicidad para Distintas Modalidades de Organización de Viajes.....	98
8.3.1	Opción A	99
8.3.2	Opción B.....	101
8.3.3	Opción C.....	101
8.3.4	Opción D	103
8.3.5	Opción E.....	104

CAPÍTULO 9 – RESUMEN Y CONSIDERACIONES FINALES..... 109

9.1	Sumario de los modelos generales obtenidos.....	109
9.2	Sumario de los modelos obtenidos para gente joven	110
9.3	Sumario de Aplicaciones Prácticas.....	110
9.3.1	Comparación de Alternativas de envío de Publicidad de Viajes con Hoteles de Lujo para Jóvenes. 111	
9.3.2	Comparación de Alternativas de envío de Publicidad de Viajes con Distintos Destinos para Jóvenes. 111	

9.3.3 Comparación de Alternativas de Publicidad para Distintas Modalidades de Organización de Viajes.....	112
9.4 Límites de la Investigación.....	113
9.5 Acerca de Futuras Investigaciones.....	113
<i>BIBLIOGRAFÍA.....</i>	<i>115</i>
<i>ANEXOS.....</i>	<i>117</i>
Anexo I. Encuesta	117
Anexo II. Técnicas estadísticas	119
Métodos Factoriales (de interdependencia)	119
Análisis de Componentes Principales (H. Hotelling).....	119
Ecuaciones estructurales	119
Análisis cluster	119
Análisis Factorial Exploratorio (Ch. Spearman, L. Thurstone).....	119
Métodos de regresión (de dependencia).....	120
Análisis Multivariante de la Varianza (Manova) (Confirmatorio) (R. Fisher).....	120
Análisis Multivariante de la Covarianza (Mancova) (Confirmatorio) (R. Fisher)	120
Análisis de Correlación Canónica (Exploratorio) (H. Hotelling).....	121
Análisis Discriminante (Confirmatorio) (R. Fisher)	121
Regresión Logística.....	121
Las Transformaciones Logit y Logística	122
La transformación “Log Odds Ratio”.....	123
La Regresión logística y los modelos Logit	123
El caso particular de la Regresión Logística Binaria.....	125
Resolviendo las Ecuaciones de Verosimilitud.....	127
Interpretación de los coeficientes de regresión.....	129
X Binaria	130
Múltiples Variables Independientes	131
Variable Dependiente Multinomial	131
Tests Estadísticos e Intervalos de Confianza.....	132
Razón de Verosimilitud y Desviación.....	133
Desviación.....	133
Test de Wald.....	135
Intervalos de Confianza.....	135
R-cuadrado	135
Diagnóstico de Residuos	136
Residuos Simples	136
Residuos de Pearson.....	137

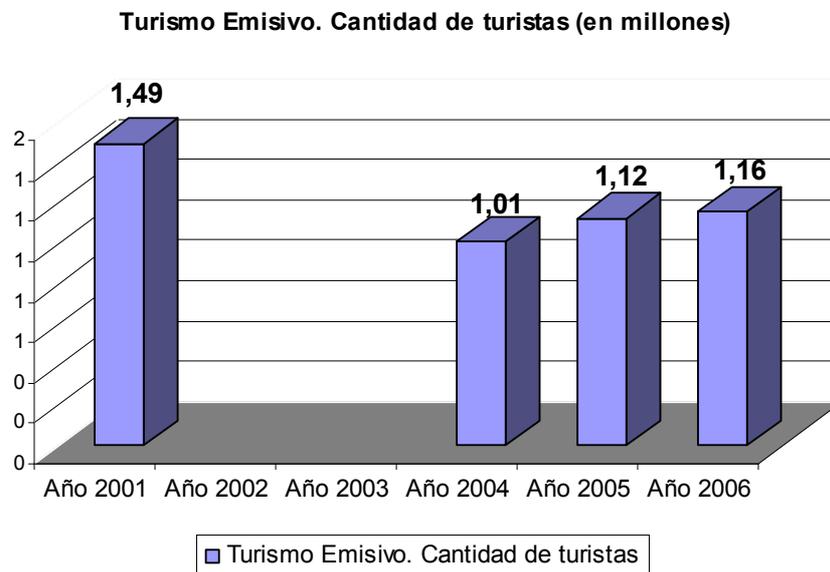
SECCIÓN I

DESARROLLO DE LA INVESTIGACIÓN

CAPÍTULO 1 - INTRODUCCIÓN

Luego de la crisis sufrida en la Argentina que estalló en Diciembre de 2001, el turismo emisor, el de los turistas residentes en la Argentina que visitan el exterior, enfrentaba un panorama oscuro. La devaluación de la moneda a principios de 2002 hacía poco tentadora la idea de enfrentar un viaje con costos en dólares. Es válido recordar que el 8 de Enero de 2002 la tasa de cambio oficial pasó de ser de un Dólar Estadounidense por cada Peso Argentino a un Dólar Estadounidense cada 1,4 Pesos Argentinos, lo que equivale, a una devaluación del 40 % con expectativas de seguir devaluando considerablemente más. Lo que hacía esta situación aún más grave, es que la mayoría de los que querían obtener Dólares a cambio de sus Pesos debían pagar aún más de la tasa oficial en mercados paralelos (hasta 2,10 Pesos por Dólar en Enero de 2002).

De esta manera, el, hasta ese momento, creciente turismo emisor se vio bruscamente frenado en el 2002. Es más, la caída fue tan notoria que hasta el primer trimestre de 2007, los valores respecto al 2001 no han sido alcanzados de nuevo. Comparando los valores del primer trimestre del 2001 y los del 2007 la caída de turistas es de un 46%¹. Para clarificar ese concepto se muestra el gráfico 1.1:



Se puede observar en el gráfico anterior una caída de un 23 % si comparamos el total del año 2001 con respecto al 2006². Sin embargo, se puede apreciar un aumento desde el año 2004 en adelante que se verá mejor a continuación:

¹ Fuente INDEC

² Los datos de 2002 y 2003 no fueron relevados por el INDEC.

Turismo Emisivo. Cantidad de turistas (en millones)

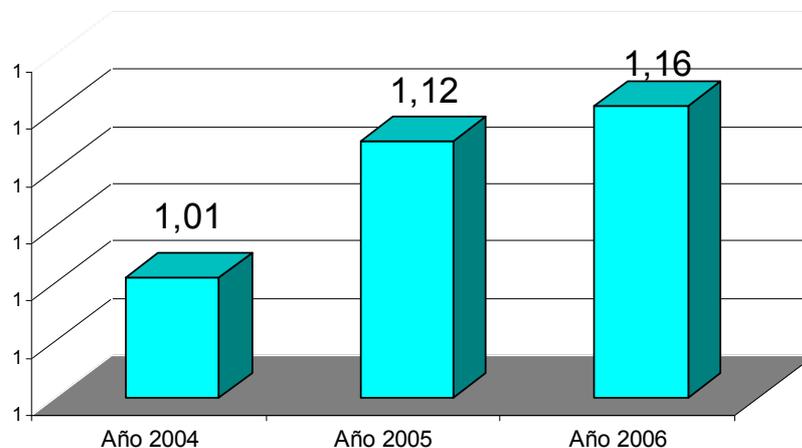


Gráfico 1.2

Turismo emisor por trimestres

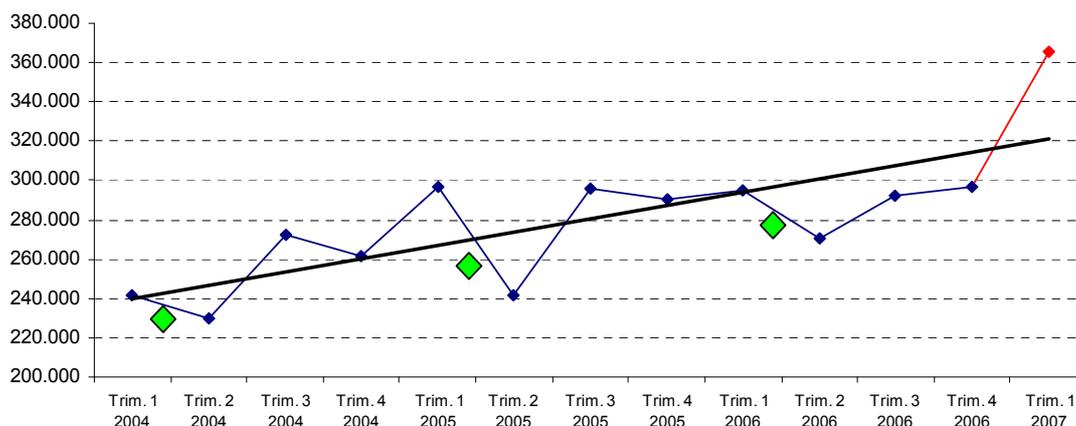


Gráfico 1.3

En los gráficos 1.2 y 1.3 se observa una tendencia alcista en la cantidad de turistas del tipo emisor. En el último podemos profundizar cuestiones como la estacionalidad (los rombos en el gráfico 1.3 muestran una mayor cantidad de turistas en los primeros que en los segundos trimestres). A su vez en el último trimestre del gráfico 1.3 (correspondiente al primer trimestre de 2007) se ve el punto máximo del período, excediendo en más de sesenta mil turistas a cualquier trimestre de 2004 en adelante.

Con respecto al desembolso de dinero que realizan los turistas, éste es indirectamente relevado en dólares norteamericanos gastados en pernoctación diaria. El gráfico 1.4 muestra la evolución de 2004 en adelante:

Dólares Promedio gastados en Pernoctación Diaria del Turismo Emisivo

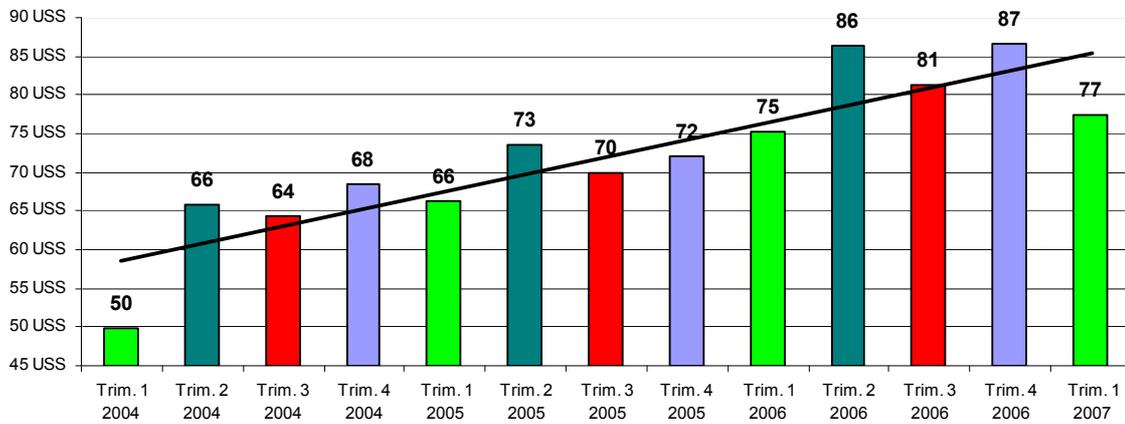


Gráfico 1.4

Del gráfico 1.4 se obtienen algunas pautas muy claras:

- La tendencia en dólares gastados por noche es alcista.
- Se observan diferencias estacionales en los distintos trimestres.
 - En los segundos trimestres se gasta más que en los primeros³
 - En los cuartos trimestres se gasta más que en los terceros.
 - En los segundos trimestres se gasta más que en los terceros.
 - Los trimestres donde más se gasta son el segundo o el cuarto

Otro dato relevante es la cantidad de días que los turistas viajan. Aquí no se dará el caso de tendencias alcistas, como hemos visto hasta ahora. Es más, la tendencia será en baja y se muestra en el gráfico 1.5:

Permanencia Promedio (días)

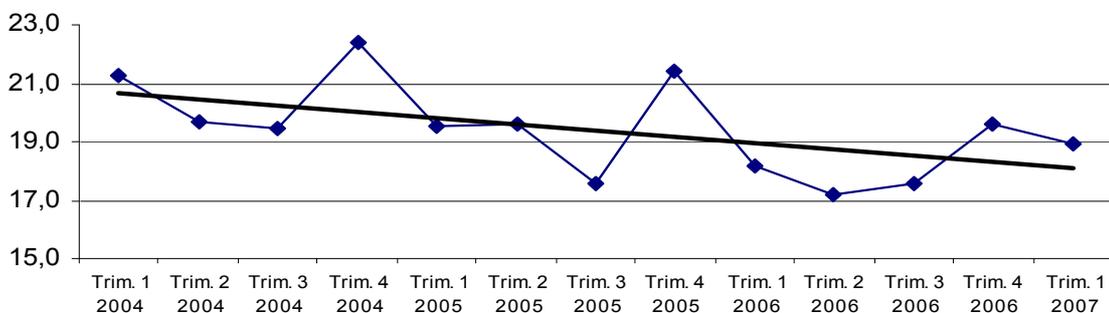


Gráfico 1.5

³ Se recuerda que no sucede lo mismo con la cantidad de turistas. Fuente INDEC.

Se observa que en los tres años relevados el de mayor permanencia promedio fue el cuarto trimestre.

Ahora bien, en estos dos últimos gráficos se vio que los dólares por día gastados venían en alza y la permanencia promedio de los viajes estaba en baja. Entonces ¿Los dólares totales gastados en cada año se encuentran en alza o en baja? Esta pregunta se hace ya que resulta fundamental para evaluar la importancia económica del negocio del turismo emisor.

El gráfico 1.6 responde la pregunta ya que surge de la multiplicación del gasto promedio por noche por la cantidad de noches por la cantidad de turistas, lo que resulta en la cantidad de dólares totales que fueron gastados año a año en pernoctación.

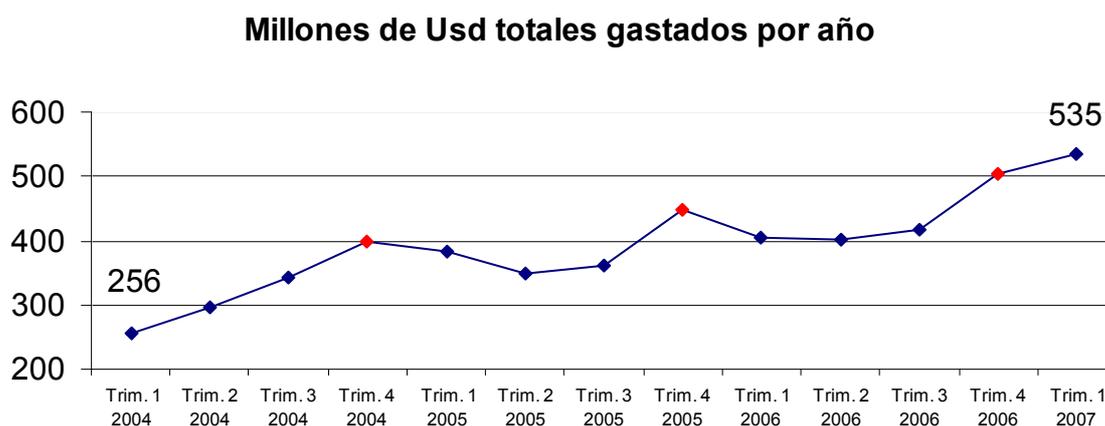
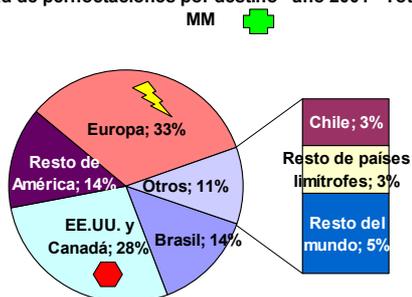


Gráfico 1.6

Se nota en el gráfico una evolución favorable ya que en sólo tres años la cantidad de dólares se incrementó en un 109 %, o sea, más del doble. Además, se observa que los cuartos trimestres son los más significativos en términos económicos.

En cuanto a la distribución de los destinos, éstos han cambiado con respecto a la época de convertibilidad. A continuación, en el gráfico 1.7, se observan dos gráficos de torta comparativos:

Cantidad de pernoctaciones por destino - año 2001 - Total 30,1



Cantidad de pernoctaciones por destino - año 2006 - Total 20,9

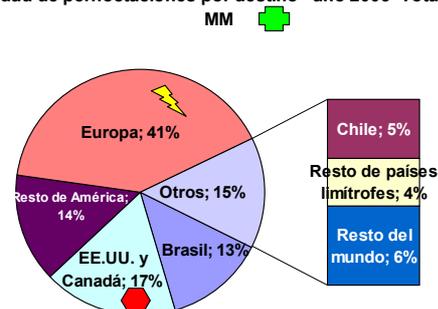


Gráfico 1.7

De esta comparación podemos extraer lo siguiente:

- La cantidad de turistas y pernoctaciones bajó considerablemente respecto de la convertibilidad.
- Los destinos EE.UU. y Canadá bajaron considerablemente como destino turístico⁴
- Europa aumentó su participación como destino turístico en un 8 %.⁵
- Los demás destinos se mantuvieron en su proporción habitual.

Con respecto al lugar de hospedaje de los turistas, el porcentaje de visitantes de hoteles de lujo⁶ se mantuvo estable. En este punto hay que aclarar que nos basamos en datos del INDEC, donde no sólo se hace foco en el turismo por placer (el cual será el centro de este trabajo) sino también en los viajes de negocios, de salud u otros motivos. Es razonable pensar que la demanda de hoteles de lujo en viajes de negocios sea más inelástica que en viajes de placer, por lo tanto como la relación viajes de negocio contra viajes de placer ha aumentado los valores de hospedaje de hoteles de lujo son lógicos. El gráfico 1.8 presenta la evolución de la utilización de hoteles de categoría.

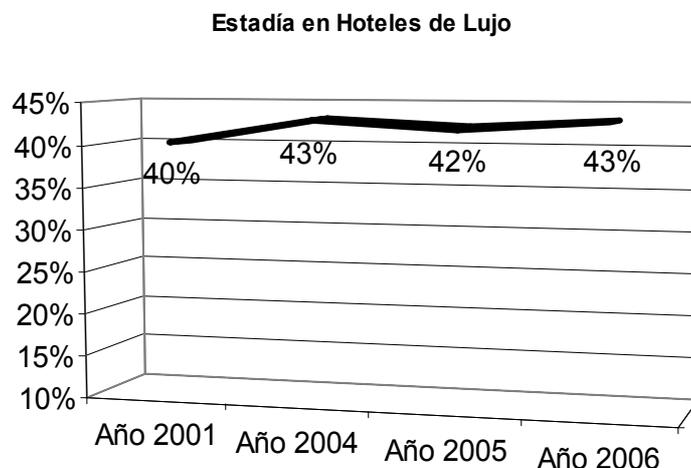


Gráfico 1.8

Para confirmar la aseveración hecha en el párrafo anterior, se muestra el gráfico 1.9:

⁴ En la "década de los 90" muchos argentinos que habitualmente no viajaban internacionalmente visitaban destinos como Orlando o Miami favorecidos por el tipo de cambio.

⁵ A pesar de haber caído la cantidad de viajantes total, se podría pensar que una pequeña proporción de viajantes pudientes siguieron visitando Europa sin importar el tipo de cambio desfavorable.

⁶ Hoteles de Lujo son los Hoteles de 4 y 5 estrellas.

Modalidad de organización del viaje y Motivo

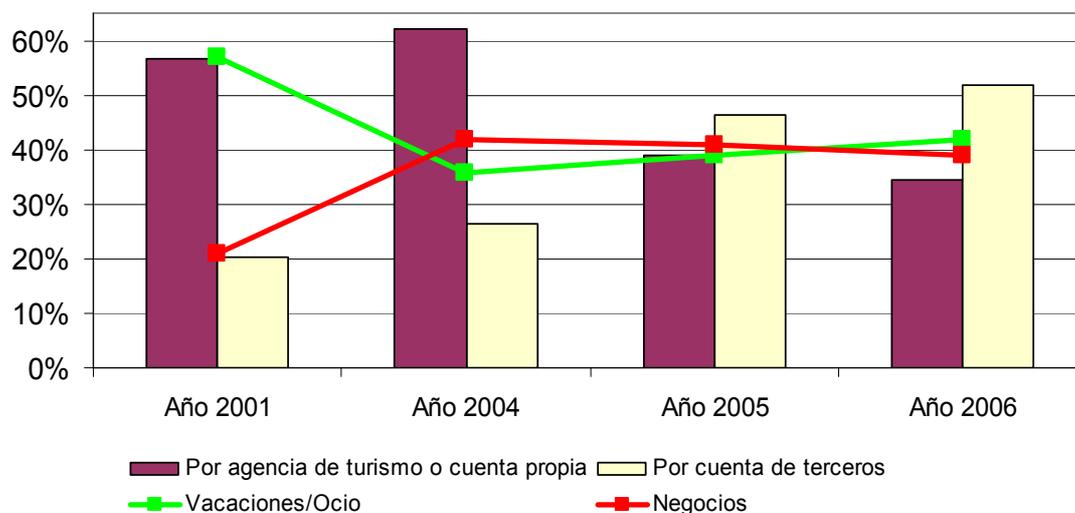


Gráfico 1.9

Como se observa en 1.7, la proporción de viajes con motivo de vacaciones u ocio bajó y por ende aumentó la proporción de viajes de negocios. Para reforzar este concepto, se tiene el aumento de viajes organizados por cuenta de terceros que son los típicos que organizan las empresas para que viajen sus empleados y gerentes.

Si bien el panorama del turismo emisor (como se vio hasta el momento), se presenta favorable y en alza, los que han subido mucho son los costos de la gestión comercial.

Cada vez los clientes tienen más posibilidades de armar sus viajes evitando las tradicionales agencias de viaje mediante el uso, por ejemplo, de Internet. Las empresas que no se adaptan a los tiempos y se quedan sólo con la promoción tradicional (gestión de vendedores, oferta del servicio por teléfono o carta) tienden a desaparecer ya que comparativamente éstas vías resultan ser más costosas.

Ante esta realidad resulta fundamental un enfoque activo en el que las empresas traten de conocer los gustos y hábitos de los compradores en potencia para adaptar sus ofertas a sus necesidades. Si se agrega a esto que hay muchas empresas en el rubro que compiten por los mismos clientes, no resulta muy eficiente realizar un enfoque de Marketing que se base en, por ejemplo, simplemente competir por precios utilizando las vías de siempre.

Es práctica usual en agencias de viaje, el envío de publicidad y promociones de viaje a toda su cartera de clientes. Si bien pueden realizar estos envíos con cierto criterio, por ejemplo, enviarle sólo a sus clientes jóvenes ofertas de paquetes turísticos para estudiantes, ¿Resulta esto eficiente? ¿Es posible descubrir ciertos atributos en los clientes que los hagan más propensos a contratar un determinado tipo de viaje, y de esta manera, ahorrar dinero no enviándole publicidad a todos los clientes?

Lógicamente, por motivos económicos, las agencias preferirían que todos sus clientes solicitaran viajes a destinos alejados y exclusivos, utilizando hoteles de

lujo. Las agencias podrían tomar como variable la situación económica de los clientes para determinar qué clientes son candidatos a estos viajes. ¿Sería esto correcto? ¿Qué otras variables influyen?

CAPÍTULO 2 - OBJETIVOS, MARCO TEÓRICO Y MODELO CONCEPTUAL

Este trabajo tiene como objetivo crear modelo predictivos que permitan dar soporte a la administración de agencias de viaje permitiendo optimizar las ofertas que éstos realizan evitando pérdidas de dinero y tiempo debido a una mala elección de clientes para un determinado paquete turístico o viceversa. De esta manera, se intentará que la aplicación del modelo permita la asignación eficiente de recursos destinados a publicidad y fuerza de ventas entre otros.

Los modelos teóricos que sirven de fundamento a este diseño de análisis son las teorías del Marketing directo y el Marketing de base de datos.⁷

Por Marketing Directo se entiende un sistema de marketing interactivo que utiliza uno o más medios publicitarios para lograr una respuesta medible y/o una transacción en cualquier lugar.

Por Marketing de Base de Datos se entiende el proceso de construir, mantener y usar bases de datos de clientes para establecer contactos y efectuar transacciones. Para que este proceso sea viable se necesita, obviamente, de una buena base de datos, que supone una colección organizada de datos exhaustivos acerca de clientes o prospectos individuales que esté actualizada y sea accesible. No alcanza, bajo ningún punto de vista, con un conjunto de nombres, direcciones y números telefónicos. Debe contener datos demográficos, psicográficos y compras anteriores entre otros.

Utilizando la información de sus bases de datos, las empresas pueden determinar con mucha mayor precisión sus mercados meta, comparando con la realización de Marketing Masivo, Marketing de Segmentos o Marketing de Nichos.

Las empresas utilizan sus bases de datos de cuatro formas:

- Para identificar prospectos
- Para decidir qué clientes deben recibir una oferta dada.
- Para hacer más profunda la lealtad a sus clientes.
- Para reactivar las compras de los clientes.

En nuestro caso, se hará mayor énfasis en las dos primeras formas.

Si se atiende a los objetivos del trabajo, resulta razonable un esquema conceptual en el que se definan modelos de dependencia, en el que haya variables dependientes que deberían ser explicadas por las variables independientes. Variables dependientes de interés podrían ser si una persona viajará como turista a Europa, si organizará sus viajes por Internet, si gastará mucho de su dinero en turismo o si no tiene planeado viajar en el futuro.

Ahora bien, se busca que las variables dependientes puedan ser explicadas por variables independientes que podrían ser por ejemplo edad de la persona, antecedentes de viajes anteriores, situación económica o estado civil.

⁷ Fuente: Dirección de Marketing, de Philip Kotler, páginas 650 a 675

Las variables dependientes e independientes de interés en nuestro caso serán explicadas en la sección siguiente.

CAPÍTULO 3 - MUESTRA

La muestra no se limita a indagar una sola cuestión en muchas personas (cómo podría ser preguntarle a un grupo de turistas si se hospedará en hoteles de lujo) o en relevar muchas cuestiones en una sola persona sino que examina muchas cuestiones relevantes acerca de viajes turísticos en muchas personas.

Para el trabajo de campo se han realizado 191 encuestas, mediante la utilización de un cuestionario. La gente encuestada comparte muchas características concernientes al alcance del proyecto. En general, se relevó a personas de clase media alta, con gran proporción de jóvenes y que comparten la característica de residir mayoritariamente en Buenos Aires. La justificación de tal elección reside en parte en el método de encuesta que se explicará en los párrafos siguientes (ya que éste hace que el procesamiento de la información se realice de forma automática evitando el ingreso de datos manuales en la PC) y también debido a que las encuestas fueron completadas mayormente por gente conocida que lo hizo desinteresadamente. No se cuenta con la infraestructura ni los recursos de grandes empresas que realizan encuestas pagas a sus clientes como pueden ser Coca-Cola, Pepsi, Marlboro u otras grandes.

Del cuestionario, que se puede consultar en el anexo I, se extrajeron las variables dependientes y las independientes que se mostrarán a continuación:

Variables Dependientes:

- ¿Planea viajar como turista dentro de los próximos 2 años, con qué destino/s lo haría?

Esta variable dependiente tiene como respuesta 13 categorías semánticas, que corresponden a agrupaciones de destinos. Se considera un horizonte de 2 años ya que más tiempo no sería de interés pues los gustos y predilecciones cambian con el tiempo. Se aclara que se puede elegir más de una respuesta. Las agrupaciones son:

1. -Chile, Bolivia o Perú
2. -Centroamérica
3. -Uruguay
4. -Brasil
5. -Europa
6. -Norteamérica (EE UU, Canadá)
7. -Oceanía o África
8. -No planeo viajar
9. -Argentina Costa
10. -Argentina Sur
11. -Argentina Norte
12. -Argentina Otro
13. -Otro País

Con estas 13 categorías se releven todos los posibles destinos a los cuales un turista puede viajar. Lógicamente, para los modelos que se obtuvieron en este proyecto se realizaron agrupaciones que se explicarán en cada caso. Resultaría

inviabile un modelo con tantas salidas o bien muy dificultoso, ya que la cantidad de encuestas para tener significatividad estadística sería enorme.

- ¿Utilizaría hoteles de 3 o más estrellas en estos futuros viajes?

Esta pregunta se contesta por sí o por no por lo que corresponde una variable dependiente dicotómica que toma el valor 1 de contestar afirmativamente y 0 de hacerlo negativamente.

- De tener pensado viajar, ¿cómo planearía su viaje?

Esta variable dependiente se contesta con 3 categorías semánticas.

- 1- Contacto con agencia de viajes
- 2- Lo planearía por Internet
- 3- Otro (Indique cómo)

En el caso que se opte por Otro se da la posibilidad de contestar cómo se planea realizar los viajes. También se da la posibilidad de marcar una o más respuestas.

Variables Independientes:

- De haber viajado como turista en los últimos 5 años, ¿con qué destino/s lo hizo?
- ¿Utilizó hoteles de 3 o más estrellas en esos viajes?
- De haber viajado, ¿cómo planeó su viaje?

Las posibles respuestas a estas tres preguntas son análogas a las que se contestan en las variables dependientes. Para el caso de la primera pregunta se consideran los 5 últimos años ya que sería a partir del período del 2002. El año 2001 fue un año muy especial en la Argentina y ya a partir del año 2002 la situación fue mejorando y pareciéndose más al horizonte actual.

Se pensó la posibilidad de preguntar los últimos 10 años pero esta tendría varios inconvenientes:

- Se estaría considerando parte del período de convertibilidad que hubo en la Argentina. Las condiciones en ese momento eran mucho más favorables para el turismo emisor y por eso no se considera correcto ni representativo para nuestro panorama presente.
- Como esta variable pretende ser explicativa de los futuros viajes que realizarán las personas, un viaje realizado hace muchos años y en condiciones totalmente diferentes a las actuales no sería representativo.
- Experiencias en Autoscoring y otros modelos de regresión logística no relevan información de larga data. Se apoyan más en eventos recientes.

- Sexo:

- 1-Masculino
- 2-Femenino

Esta es una variable dicotómica que tomará el valor 1 de ser masculino y 0 de responder femenino.

Estado Civil:

- 1-Soltero/a, divorciado/a o viudo
- 2-Casado/a o en pareja
- 3-Casado/a con hijos/as menores a 18 años

Esta variable tiene tres categorías semánticas. Se agrupó de esta manera ya que a priori se podría pensar que resultaría menos probable realizar cierto tipos de viajes si una persona está casada con hijos que no puede dejar solos, comparando con una persona que tiene más libertad de acción en ese sentido pues está sola.

Edad:

- 1-15 a 30
- 2-30 a 45
- 3-45 a 60
- 4-Más de 60.

La edad se agrupa en cuatro categorías numéricas. Se busca la comodidad del que contesta pues se marca la correcta con una cruz y además porque pueden haber encuestados que se sientan incómodos revelando su edad real. De esta manera, obtenemos información que nos es relevante que no obtendríamos si preguntáramos la edad real porque el encuestado podría abstenerse a contestar o lo que es peor: mentir.

Del dinero disponible para gastar en cosas que le gustan. ¿Cuánto destina a viajes turísticos?

- 1-Poco.
- 2-Medio
- 3-Mucho

Se utilizaron estas tres categorías semánticas posicionándose sobre una escala de importancia ya que hubiera resultado muy difícil para el encuestado colocar una cifra o un porcentaje de sus ingresos. Se recuerda que no todos los encuestados trabajan o tienen ingresos y que gran parte de los trabajadores no tienen ingresos fijos. Adicionalmente, estimar un porcentaje de los ingresos destinados al turismo requeriría de cálculo y esfuerzo mental en el encuestado que seguramente muchos no estarían dispuestos a realizar o bien se equivocarían al hacerlo.

¿Cómo considera su situación económica?

- 1-Mala (no podría/no me podrían pagar un pequeño viaje con destino local)
- 2-Regular (Sólo podría/me podrían pagar un viaje con destino local pero pocos días)
- 3-Buena (Podría/me podrían pagar un viaje a un destino local un par de semanas o pagar un viaje al exterior pero no más de una semana)
- 4-Muy Buena (Podría/ me podrían pagar un viaje al exterior de un par de semanas o más)

En este caso se optó por posicionarse sobre una escala de calificación de cuatro categorías. Sería ideal tener datos objetivos acerca de este asunto pero en una encuesta eso resulta muy difícil. Mucha gente se rehusaría a proveer esta información ya que es algo privado y que generalmente no se comenta. Obsérvese que a cada una de las opciones se les colocó un comentario para dar referencias acerca de qué se considera cada situación económica. Por ejemplo, las personas que sólo podrían pagar un pequeño viaje local debían contestar malo. En este sentido se nota el alcance del proyecto ya que una situación económica mala en este país sería de indigencia, pero se recuerda al público de clase media alta al que fue dirigida la encuesta.

Otra situación se da para la gente que no tiene ingresos, los que no trabajan. ¿Cómo definen éstos su situación económica? Esto aplica por ejemplo a jóvenes del primer grupo de edad (entre 15 y 30). Por eso se aclaró en cada opción si podría o le podrían pagar tal o cual viaje.

□ Estudios:

1-Secundario

2-Universitario incompleto/ en curso

3-Universitario completo

En este caso se podrían haber colocado más opciones como puede ser Primario incompleto, Primario completo o Secundario incompleto. Debido al target de la encuesta, se descartó la opción de estudios Primarios ya que ésta se cumple inexorablemente en todos los casos. También se podría haber incluido la opción estudios Terciarios pero para simplificar, no se hizo. Queda a criterio del encuestado si se ubica en Universitarios Incompletos o Completos en este caso.

□ ¿Trabaja?

1-Si
2-No

Esta es una variable dicotómica que tomará el valor 1 de ser afirmativa la respuesta y 0 si se responde que no.

Una vez definida la muestra, la recogida de datos se hizo por el método de distribución de encuestas, tanto impresas como por vía e-mail, y también utilizando las facilidades de la página Web www.encuestafacil.com que se explicará sucintamente a continuación:

Encuestafacil.com es una herramienta Web de encuestas online utilizada en España y Latinoamérica. Permite a los usuarios elaborar por si mismos, de una forma rápida y sencilla, encuestas que ayuden en la toma de decisiones. Permite obtener información en poco tiempo y destinando pocos recursos.

Con respecto al target al que va dirigida esta herramienta se encuentran todo tipo de empresas, con independencia de su tamaño o sector de actividad, organismos públicos, profesionales de la educación, instituciones académicas y estudiantes, ONGs., investigadores independientes, Institutos de investigación y consultores que quieran utilizar esta plataforma para cubrir la fase de campo on-line.

La metodología consta de tres pasos: la creación, la recopilación y el análisis de los resultados.

En cuanto a la creación: se puede hacer una encuesta desde cero, desde plantillas predeterminadas o desde encuestas ya realizadas por el usuario.

Hay ejemplos muy claros para facilitar esta tarea. En nuestro caso se optó por crear la encuesta desde cero.

En cuanto a la recopilación: se pueden enviar las invitaciones por Email desde una dirección de correo propia o desde EncuestaFacil.com. Este método permite hacer un seguimiento de las personas que responden, ya que es el único método que registra la identidad de los encuestados. Si no se puede optar por un Link en una página Web o una encuesta en pop-up. En nuestro caso se optó por enviar mails desde una dirección propia.

En cuanto al análisis: se presentan cuadros de frecuencia, se da la posibilidad de colocar filtros en los resultados, ver informes de resultados, resúmenes y la descarga de resultados en formato CSV, utilizable por Excel, SPSS y programas similares. Esta última fue la utilizada en este proyecto.

Una vez recogidas las encuestas, se tuvo que realizar una depuración debido a errores al contestar las encuestas impresas. El error más común era contestar en el destino que se había visitado o se iba a visitar de esta manera:

4. De haber viajado como **turista** en los últimos 5 años, ¿con qué destino/s lo hizo? (Marque una o más respuestas)

- Destinos:
 -Chile, Bolivia o Perú
 -Centroamérica
 -Uruguay
 -Brasil
 -Europa
 -Norteamérica (EE UU, Canadá)
 -Oceanía o África
 -No viajé
 -X Argentina (**Indique Costa, Sur, Norte, etc.**) u otro (Indique país):

Como se ve arriba, era muy común que se contestara con una cruz: Argentina, sin indicar las opciones dentro de ésta, cómo se exigía en negrita. Se aclara que esto sucedió sólo con las encuestas escritas ya que el diseño de encuestafacil.com se construyó posteriormente al impreso mejorando las fallas que presentaba éste.

De esta manera, la muestra final con la que se ha realizado nuestro análisis cuenta con 191 casos. Este tamaño de muestra es suficiente para asegurar un análisis fiable de la técnica estadística aplicada en este estudio (que se justificará y definirá en el capítulo 4). Esta técnica recomienda una razón mínima de diez o quince casos por cada variable independiente. Puesto que las variables predictoras de nuestro modelo son diez, entonces el número de casos mínimo que asegura la fiabilidad del análisis debería ser de cien, cifra superada con la muestra actual.

Los resultados globales de la encuesta se muestran en los gráficos 3.1 al 3.4.

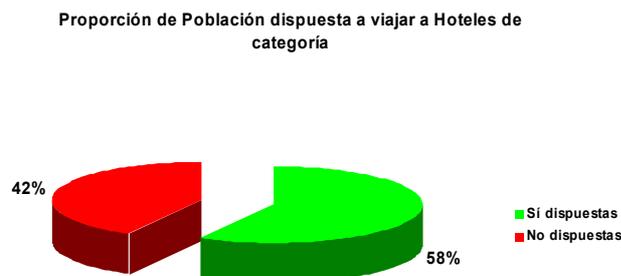


Gráfico 3.1

Regiones Máximas a las que Planean Viajar los Encuestados

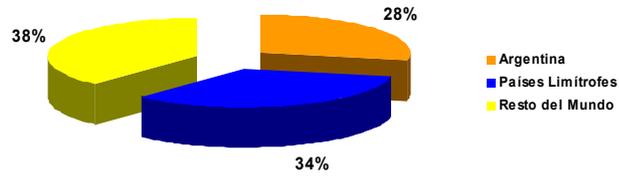


Gráfico 3.2

Proporción de Población dispuesta a Viajar a Europa

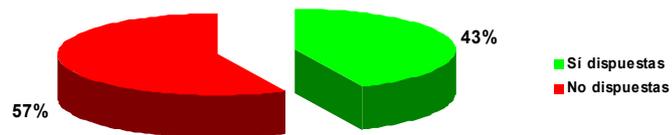


Gráfico 3.3

Modalidades de Organización de Viajes Únicas

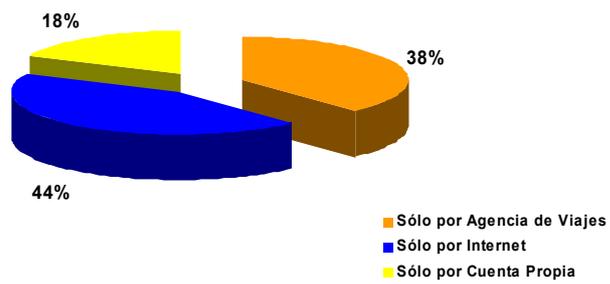


Gráfico 3.4

CAPÍTULO 4 - PROCEDIMIENTOS ESTADÍSTICOS

El esquema teórico que se plantea nos conduce a aplicar un análisis estadístico de regresión por cuanto que *a priori* conocemos cuáles son las variables dependientes y cuáles las independientes. Por tanto, se excluyen técnicas estadísticas a aplicar para análisis de interdependencia, tales como el análisis de componentes principales, las ecuaciones estructurales, el análisis cluster o el análisis factorial.

Dentro de las técnicas estadísticas de aplicación en análisis de regresión (dependencia) quedan descartados el análisis multivariante de la varianza, el análisis multivariante de la covarianza y la correlación canónica, puesto que los modelos presentan sólo una variable dependiente y no varias, como es el caso para el que están prescritas aquellas.

Dentro de los posibles tratamientos estadísticos de aplicación en análisis de dependencia con una variable dependiente, las técnicas más apropiadas son el análisis discriminante y los modelos de regresión logística, por cuanto son los únicos que permiten abordar modelos en los que existe una variable dependiente categórica. De estas dos posibilidades, se recurre a la regresión logística tanto binaria como multinomial puesto que mediante éstas se puede predecir la probabilidad de ocurrencia de un evento a partir de los valores que presentan una serie de variables independientes (que pueden ser categóricas como se presenta en nuestro caso). Sin embargo, también se presentarán los resultados del análisis discriminante para complementar y comparar el análisis.

En el anexo II se explicarán sucintamente todas las técnicas mencionadas y en detalle la técnica de regresión logística.

CAPÍTULO 5 - PASOS PARA LA OBTENCIÓN DE UN MODELO DE REGRESIÓN LOGÍSTICA BINARIA

En esta sección se explica el procedimiento realizado para obtener un modelo de Regresión Logística Binaria. La regresión Binaria es un caso particular (el más simple) de la Regresión Logística Multinomial. Se explica por separado ya que presenta la ventaja de poder ser resuelto en forma integral con la ayuda del popular software MS Excel, cosa que no sucede con los modelos más complejos. La explicación de la obtención del modelo será descripta siguiendo un ejemplo.

Consideremos un Doctor en Medicina que quiere estimar la probabilidad de que un paciente sufra la Rotura de Ligamentos Cruzados en una rodilla. Para esto, realiza un estudio en una lista de 44 pacientes durante un año determinado.

En este caso podríamos pensar, por ejemplo, tres variables explicativas: Kilogramos de Sobrepeso, Horas de Entrenamiento Semanal y si tiene Lesiones Previas.

La variable Kilogramos de Sobrepeso podrá tomar los valores 0 (de no presentar sobrepeso) a números mayor a 0. Por ejemplo, una persona con cinco kilogramos de sobrepeso tomará el valor 5 en esta variable.

La variable Horas de Entrenamiento semanal podrá tomar los valores 0 (de no entrenar en la semana) a números mayores a 0. Por ejemplo, una persona que entrena cuatro horas por semana tomará el valor 4 en esta variable.

La variable Tiene Lesiones Previas es binaria y tomará el valor 1 de presentar lesiones previas y 0 de no haber sufrido lesión alguna.

La variable dependiente en este caso (Rotura de Ligamentos) presentará el valor 1 de haber sufrido el paciente esta lesión durante el año de análisis y 0 de no haber ocurrido.

5.1 Estimación de los Parámetros por Máxima Verosimilitud

Como se mencionará en el anexo II se debe maximizar el logaritmo natural de la función de similitud:

$$Ln(l) = \sum_{t=1}^{44} \left| 1 - Y_t - \frac{1}{1 + e^{(-\beta_0 + \beta_1 X_{1t} + \beta_2 X_{2t} + \beta_3 X_{3t})}} \right|$$

Fórmula 5.1

Ahora sí estamos en condiciones de maximizar utilizando el complemento Solver el logaritmo natural de la función de verosimilitud.

Se muestra a continuación una vista de esto en el gráfico 5.1:

La *celda objetivo* corresponde a la celda donde se encuentra el logaritmo natural de la función de verosimilitud (ln L).

En *valor de la celda objetivo* se coloca la opción máximo (ya que se pretende maximizar).



Gráfico 5.1

Cambiando las celdas hace referencias a las celdas donde se encuentran los valores de b_0 , b_1 , b_2 y b_3 que son los valores semilla que calculamos anteriormente y están siendo utilizados en las fórmulas de cálculo para modelar la Y y lógicamente para la función de verosimilitud.

Sujeto a las siguientes restricciones⁹ es un recaudo que se toma para obtener una solución convergente. En este caso se busca que:

$$e^{-(b_0 + b_1 X_1 + b_2 X_2 + b_3 X_3)}$$

Fórmula 5.4

no tome un valor tan grande que $e^{-(b_0 + b_1 X_1 + b_2 X_2 + b_3 X_3)}$ la solución no converja. Nótese igual que el valor de restricción (50000000) es un número muy grande. Por lo tanto, al ser poco restrictivo, no invalida el modelo.

⁹ Artificio explicado en las clases de Estadística Aplicada y Estadística Superior de Roberto M. García.

Los valores de los coeficientes finales obtenidos con Solver son:

(b)		Coeficientes	Variable
b0	Intercepción	5,75	Intercepción
b1	Variable X 1	1,57	Kilos de Sobrepeso
b2	Variable X 2	-7,82	Horas de Entrenamiento Semanal
b3	Variable X 3	10,41	Tuvo lesiones anteriores?

Tabla 5.2

Nótese como los valores b1 y b2 arrojan valores positivos. Esto es lógico ya que es de esperar que al aumentar el sobrepeso o de ya haber sufrido lesiones las personas sean más propensas a lesionarse. Lo contrario ocurre con las horas de entrenamiento. Al aumentar éstas disminuirán las chances de sufrir roturas. Véase que el valor b2 es negativo.

La ecuación de estimación de probabilidad de rotura final será.

$$Y = \frac{1}{1 + e^{-(5,75 + 1,57 X_1 - 7,82 X_2 + 10,41 X_3)}}$$

Fórmula 5.5

Supongamos una persona que:

- No presenta sobrepeso: b1 = 0
- Entrena una hora por semana: b2 = 1
- No ha sufrido lesiones: b3= 0

Utilizando la ecuación de Y, la probabilidad que arroja el modelo de que este paciente sufra una rotura de ligamentos es: 0,11 (11 %).

5.2 Validación del Modelo

Para la validación global del modelo, el paso siguiente es colocar los valores semilla 0 en b1, b2 y b3 y volver a correr el Solver sólo modificando los valores de b0 en *cambiando las celdas*. De esta manera, se obtiene el estimador maxiverosímil de b0 (b0 = -0,76), y que maximiza el logaritmo de la verosimilitud, que resulta Ln Lo = -27,52, con lo cual se obtiene el estadístico 2(In L-In Lo) = 48,76, que le corresponde una probabilidad acumulada derecha 1,46.10⁻¹⁰, para un Chi-Cuadrado de tres grados de libertad (grados de libertad iguales al número de variables independientes). La hipótesis nula β1= β2= β3=0 se rechaza con certeza práctica y se valida globalmente el modelo.

Se muestra una imagen de la pantalla de MS Excel en este caso:

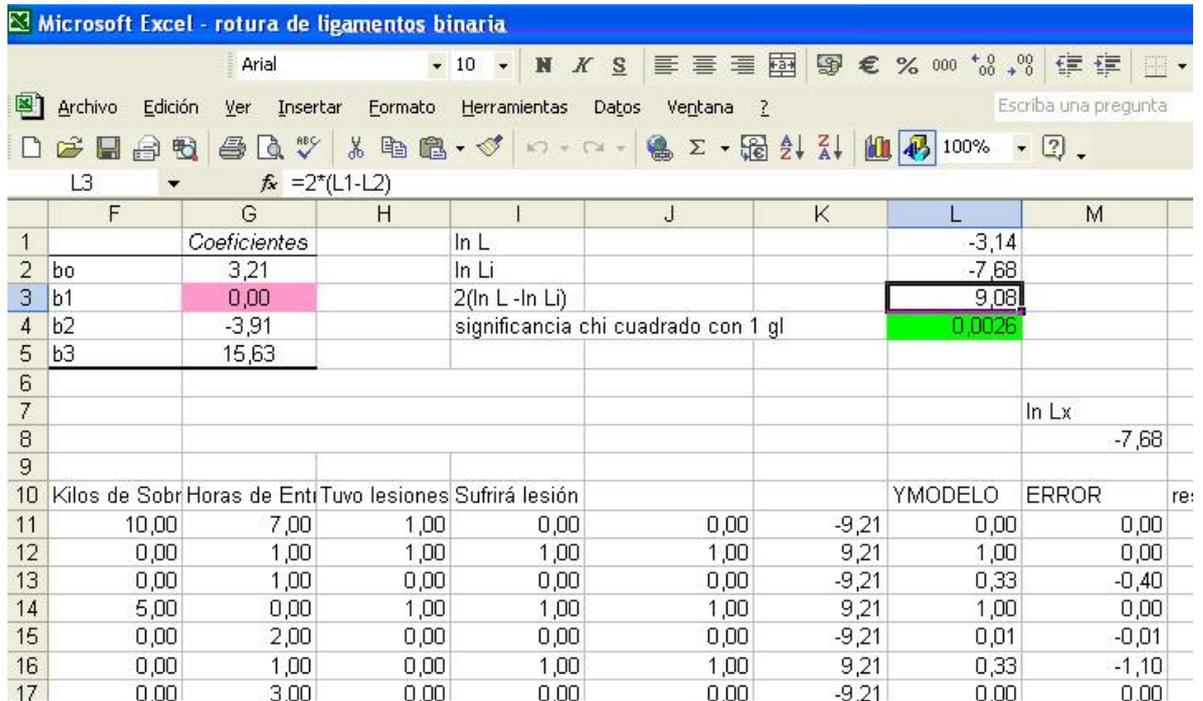
	F	G	H	I	J	K	L	M
1		Coeficientes		In L				-3,14
2	bo	-0,76		In LO				-27,52
3	b1	0,00		2(In L -In Lo)				48,76
4	b2	0,00		significancia chi cuadrado con 3 gl			0,000000000146	
5	b3	0,00						
6								In Lx
7								
8								
9								
10	Kilos de Sobr	Horas de Enti	Tuvo lesiones	Sufrirá lesión			YMODELO	ERROR
11	10,00	7,00	1,00	0,00	0,00	-9,21	0,32	
12	0,00	1,00	1,00	1,00	1,00	9,21	0,32	
13	0,00	1,00	0,00	0,00	0,00	-9,21	0,32	
14	5,00	0,00	1,00	1,00	1,00	9,21	0,32	
15	0,00	2,00	0,00	0,00	0,00	-9,21	0,32	
16	0,00	1,00	0,00	1,00	1,00	9,21	0,32	
17	0,00	3,00	0,00	0,00	0,00	-9,21	0,32	

Gráfico 5.2

Lo que sigue es evaluar la significancia individual de los coeficientes. Para el caso de $\beta_1 = 0$, lo que se hace es colocar los valores semillas originales salvo al valor de b_1 que se le coloca 0. En este caso se maximiza como ya hemos visto pero en *cambiando las celdas* sólo se colocan las celdas correspondientes a b_0 , b_2 y b_3 . A su vez en *Sujeto a las siguientes restricciones* se modificó el valor original por uno más pequeño ya que el modelo no convergía.

La hipótesis nula $\beta_1=0$ se rechaza ya que al estadístico $2(\ln L - \ln L_i) = 9,08$ le corresponde una probabilidad acumulada derecha de 0,0026, para un Chi-Cuadrado de 1 grado de libertad. La significación es aún más fuerte ya que 0,0026 es el valor para el ensayo bilateral y como el ensayo es unilateral debido a que conocemos el sentido del coeficiente la significación será de 0,0013. Por lo tanto, no se procede a retirar la variable X_1 del análisis.

Se muestra una imagen de la pantalla de MS Excel en este caso en el gráfico 5.3:



	F	G	H	I	J	K	L	M
1		<i>Coefficientes</i>		ln L			-3,14	
2	bo	3,21		ln Li			-7,68	
3	b1	0,00		2(ln L - ln Li)			9,08	
4	b2	-3,91		significancia chi cuadrado con 1 gl			0,0026	
5	b3	15,63						
6								
7								ln Lx
8								-7,68
9								
10	Kilos de Sobr	Horas de Enti	Tuvo lesiones	Sufrirá lesión			YMODELO	ERROR
11	10,00	7,00	1,00	0,00	0,00	-9,21	0,00	0,00
12	0,00	1,00	1,00	1,00	1,00	9,21	1,00	0,00
13	0,00	1,00	0,00	0,00	0,00	-9,21	0,33	-0,40
14	5,00	0,00	1,00	1,00	1,00	9,21	1,00	0,00
15	0,00	2,00	0,00	0,00	0,00	-9,21	0,01	-0,01
16	0,00	1,00	0,00	1,00	1,00	9,21	0,33	-1,10
17	0,00	3,00	0,00	0,00	0,00	-9,21	0,00	0,00

Gráfico 5.3

Análogamente, se realizó el mismo procedimiento para las 2 variables independientes restantes. Las dos resultaron significativas. De no haberlo sido, si le hubiera correspondido un valor de probabilidad acumulada derecha de más de 0,1, para un Chi-Cuadrado de 1 grado de libertad, se hubiera procedido a retirar la variable del análisis y evaluar nuevamente el modelo sin ésta.

CAPÍTULO 6 - PASOS PARA LA OBTENCIÓN DE UN MODELO DE REGRESIÓN LOGÍSTICA MULTINOMIAL

En esta sección se explica el procedimiento realizado para obtener un modelo de Regresión Logística Multinomial. Para hacer la explicación más comprensible se desarrollarán los pasos siguiendo un ejemplo.

El ejemplo consiste en una encuesta en la que los participantes responden acerca de ciertos aspectos personales y su género de película preferido. Usando regresión logística multinomial, se pueden crear perfiles de personas que son más propensas a estar interesadas en determinado género con el fin de poder diseñar una publicidad acorde a este perfil de persona.

En la encuesta se releva la edad, el sexo, el estado civil y la situación laboral de las personas.

En cuanto a la edad, se establecen cuatro categorías:

- Categoría 1: Menores a 30 años. La variable independiente toma el valor 1.
- Categoría 2: Entre 31 y 45 años. La variable independiente toma el valor 2.
- Categoría 3: Entre 46 y 60 años. La variable independiente toma el valor 3.
- Categoría 4: Mayor a 60 años. La variable independiente toma el valor 4.

En cuanto al sexo, se establecen dos categorías:

- Categoría 1: Sexo Femenino. La variable independiente toma el valor 0.
- Categoría 2: Sexo Masculino. La variable independiente toma el valor 1.

En cuanto a la situación laboral, se establecen dos categorías:

- Categoría 1: Desocupado. La variable independiente toma el valor 0.
- Categoría 2: Ocupado. La variable independiente toma el valor 1.

Cada una de estas elegirá uno de los tres géneros de película que consideren su favorita. Si la película preferida es:

- Drama: La variable dependiente toma el valor 1
- Comedia: La variable dependiente toma el valor 2
- Acción: La variable dependiente toma el valor 3

El Software utilizado en este caso es SPSS v13.0, pues resulta más sencilla la solución del modelo comparando con la utilización del MS Excel.

6.1 Indagaciones Exploratorias previas

6.1.1 Análisis de Correlaciones Bivariadas¹⁰

Se intenta obtener un modelo explicativo parsimonioso que incluya un número de variables predictoras razonables. De este modo, resulta necesario asegurarse previamente de que las variables independientes incluidas en el modelo sean relevantes en la predicción de la variable dependiente. Para lograr

¹⁰ En el caso de utilizar SPSS, en la barra de herramientas utilizar: Analizar/Correlaciones/Bivariadas.

esto se realiza el análisis de correlaciones bivariantes a partir de la matriz de correlaciones que incluye a la variable dependiente a analizar y a todas las variables independientes.

Ejecutado el análisis se desprenden las variables independientes que resultan significativas al nivel 0,05 bilateral.

De estos indicadores se concluye que el resto de las variables independientes son irrelevantes a la hora de explicar la variable dependiente, por lo que se considera oportuno excluirlas del análisis.

En el caso de nuestro ejemplo:

		Correlaciones				
		Género de Película Preferido	Categoría de Edad	Sexo	Estado Civil	Situación Laboral
Género de Película Preferido	Correlación de Pearson	1	,083*	-,030	-,112**	-,148**
	Sig. (bilateral)		,013	,369	,001	,000
	N	880	880	880	880	880
Categoría de Edad	Correlación de Pearson	,083*	1	-,019	-,029	-,230**
	Sig. (bilateral)	,013		,567	,384	,000
	N	880	880	880	880	880
Sexo	Correlación de Pearson	-,030	-,019	1	,091**	,035
	Sig. (bilateral)	,369	,567		,007	,303
	N	880	880	880	880	880
Estado Civil	Correlación de Pearson	-,112**	-,029	,091**	1	,020
	Sig. (bilateral)	,001	,384	,007		,550
	N	880	880	880	880	880
Situación Laboral	Correlación de Pearson	-,148**	-,230**	,035	,020	1
	Sig. (bilateral)	,000	,000	,303	,550	
	N	880	880	880	880	880

*. La correlación es significativa al nivel 0,05 (bilateral).

** . La correlación es significativa al nivel 0,01 (bilateral).

Tabla 6.1

La variable sexo no resulta significativa con respecto al género de película preferido pues presenta un valor de significatividad mayor a 0,05 (0,369) por lo que se excluye del análisis.

6.1.2 Análisis de Contrastes de Razón de Verosimilitud¹¹

El siguiente paso consiste en determinar la relevancia de las variables que han superado la prueba anterior, lo que se ha de realizar son contrastes de razón de verosimilitud para comprobar si la eliminación de alguna de estas variables mejora el ajuste del modelo.

Se llega a la conclusión que las variables a considerar luego de estos contrastes son las que tienen una significatividad menor o igual a 0,05.

En el caso de nuestro ejemplo:

¹¹ En el caso de utilizar SPSS, en la barra de herramientas ir a: Analizar/Regresión/Logística Multinomial. Se recomienda que en *estadísticos* se seleccionen todas las opciones disponibles que figuran bajo el título *modelo*.

Contrastes de la razón de verosimilitud

Efecto	Criterio de ajuste del modelo	Contrastes de la razón de verosimilitud		
	-2 log verosimilitud del modelo reducido	Chi-cuadrado	gl	Sig.
Intersección	125,528 ^a	,000	0	.
catdeedad	412,803	287,275	6	,000
estadciv	152,875	27,347	2	,000
sitlab	149,386	23,859	2	,000

El estadístico de chi-cuadrado es la diferencia en las -2 log verosimilitudes entre el modelo final y el modelo reducido. El modelo reducido se forma omitiendo un efecto del modelo final. La hipótesis nula es que todos los parámetros de ese efecto son 0.

- a. Este modelo reducido es equivalente al modelo final ya que la omisión del efecto no incrementa los grados de libertad.

Tabla 6.2

Por lo tanto, en este caso, se consideran todas las variables ya que presentan todos valores de significatividad menor a 0,05.

6.1.3 Evaluar Ausencia de Colinealidad entre las Variables Independientes

El siguiente paso consiste en corroborar la ausencia de colinealidad entre las variables independientes que han sido preseleccionadas en la instancia anterior. Para eso se utiliza nuevamente la matriz de correlaciones de 6.1.1 y se verifica que los coeficientes de correlación de Pearson entre estas variables no sea superior a 0,6, lo que denotaría una existencia de importante colinealidad. De no haber evidencia que denote este problema no se procede a eliminar ninguna variable. De haber evidencia, se procederá criteriosamente a quitar alguna de las variables que presenten colinealidad.

En el ejemplo:

Correlaciones

		Categoría de Edad	Estado Civil	Situación Laboral
Categoría de Edad	Correlación de Pearson	1	-,029	-,230**
	Sig. (bilateral)		,384	,000
	N	880	880	880
Estado Civil	Correlación de Pearson	-,029	1	,020
	Sig. (bilateral)	,384		,550
	N	880	880	880
Situación Laboral	Correlación de Pearson	-,230**	,020	1
	Sig. (bilateral)	,000	,550	
	N	880	880	880

** . La correlación es significativa al nivel 0,01 (bilateral).

Tabla 6.3

Las correlaciones de Pearson entre las variables explicativas no exceden el valor de 0,6, por lo tanto no se elimina ninguna variable del análisis.

6.2 Planteo de Hipótesis

Como este trabajo pretende elaborar modelos predictivos con el fin de aportar herramientas para la optimización del área comercial en el sector turismo, no sólo se necesita determinar que variables son las que determinan, por ejemplo, si un turista estaría dispuesto a concurrir a un hotel de lujo sino también cual es la relación (positiva o negativa) entre éstas y la variable de salida. En otras palabras se pretende responder: dado un valor de una cierta variable explicativa: ¿Aumenta o disminuye la probabilidad de ocurrencia de la variable explicada?

Para todas las variables independientes que han superado las pruebas anteriores se plantean hipótesis *a priori* de cómo afectarán y si aumentarán o disminuirán la probabilidad de ocurrencia de la variable a explicar.

En el caso de nuestro ejemplo, se podría plantear la hipótesis de que a mayor edad de las personas estas perderían cierto interés en las películas de acción y se podrían inclinar más a las comedias.

6.3 Resultados del modelo

6.3.1 Estimación de Los Coeficientes y Evaluación del Modelo Obtenido.

Luego de la ejecución de un análisis de regresión logística multinomial mediante algún Software (como puede ser SPSS 13.0) se deben hacer estudios para corroborar la validez del modelo.

*Lo primero que se debe verificar es la información del ajuste del modelo. Se debe verificar que el Test de Razón de Verosimilitud que se observa en la Tabla de

Información del ajuste del modelo arroje un valor de significancia menor a 0,05. Esto indica que el modelo final supera al modelo nulo (el que no contempla ninguna variable).

En el caso de nuestro ejemplo:

Información del ajuste del modelo

Modelo	Criterio de ajuste del modelo	Contrastes de la razón de verosimilitud		
	-2 log verosimilitud	Chi-cuadrado	gl	Sig.
Sólo la intersección	490,197			
Final	125,528	364,669	10	,000

Tabla 6.4

En este caso el valor de significatividad es inferior a 0,05 por lo que el modelo final supera al modelo nulo y se procede con el análisis.

*Luego se debe comprobar la bondad del ajuste del modelo. Para esto se utilizan los estudios de Pearson y Desviación. Valores de significatividad mayores a 0,1 indican que los datos son consistentes con los supuestos del modelo.

En el caso de nuestro ejemplo:

Bondad de ajuste

	Chi-cuadrado	gl	Sig.
Pearson	9,951	20	,969
Desviación	10,243	20	,964

Tabla 6.5

Los valores de significatividad son mayores a 0,1 por lo que se procede con el análisis.

*El próximo paso consiste en verificar la contribución de cada efecto al modelo. Si los valores de significancia de las variables independientes en la Tabla de Contrastes de la Razón de Verosimilitud son menores a 0,05 entonces los efectos contribuyen al modelo.

En el caso de nuestro ejemplo:

Contrastes de la razón de verosimilitud

Efecto	Criterio de ajuste del modelo	Contrastes de la razón de verosimilitud		
	-2 log verosimilitud del modelo reducido	Chi-cuadrado	gl	Sig.
Intersección	125,528 ^a	,000	0	.
catdeedad	412,803	287,275	6	,000
estadciv	152,875	27,347	2	,000
sitlab	149,386	23,859	2	,000

El estadístico de chi-cuadrado es la diferencia en las -2 log verosimilitudes entre el modelo final y el modelo reducido. El modelo reducido se forma omitiendo un efecto del modelo final. La hipótesis nula es que todos los parámetros de ese efecto son 0.

- a. Este modelo reducido es equivalente al modelo final ya que la omisión del efecto no incrementa los grados de libertad.

Tabla 6.6

Como los valores de significatividad son menores a 0,05 se puede decir que los efectos contribuyen al modelo.

*Posteriormente, se procede a verificar la significatividad que se obtiene con los tests de Wald en la Tabla de Estimados de parámetros. Si todos los valores de significatividad son menores a 0,05 los parámetros se consideran distintos de cero. Esto dará como consecuencia que los intervalos de confianza de 95 % de los e^B no contengan el valor 1 ya que eso denotaría la probabilidad de que esa categoría no tenga una respuesta distinta que la categoría de referencia.

En el caso de nuestro ejemplo:

Estimaciones de los parámetros

Género de Película Preferido		B	Error típ.	Wald	gl	Sig.	Exp(B)	Intervalo de confianza al 95% para Exp(B)	
								Límite inferior	Límite superior
Drama	Intersección	-,421	,301	1,958	1	,162			
	[catedeedad=1]	1,016	,318	10,210	1	,001	2,761	1,481	5,147
	[catedeedad=2]	1,363	,323	17,762	1	,000	3,907	2,073	7,363
	[catedeedad=3]	,714	,341	4,380	1	,036	2,043	1,046	3,987
	[catedeedad=4]	0 ^b	.	.	0
	[estadciv=0]	-,755	,194	15,115	1	,000	,470	,321	,688
	[estadciv=1]	0 ^b	.	.	0
	[sitlab=0]	-,772	,184	17,557	1	,000	,462	,322	,663
	[sitlab=1]	0 ^b	.	.	0
Comedia	Intersección	,828	,224	13,618	1	,000			
	[catedeedad=1]	-3,578	,401	79,503	1	,000	,028	,013	,061
	[catedeedad=2]	-2,225	,267	69,578	1	,000	,108	,064	,182
	[catedeedad=3]	-1,041	,222	22,076	1	,000	,353	,229	,545
	[catedeedad=4]	0 ^b	.	.	0
	[estadciv=0]	,418	,208	4,043	1	,044	1,519	1,011	2,284
	[estadciv=1]	0 ^b	.	.	0
	[sitlab=0]	,162	,184	,775	1	,379	1,175	,820	1,685
	[sitlab=1]	0 ^b	.	.	0

a. La categoría de referencia es: Acción.

b. Este parámetro se ha establecido a cero porque es redundante.

Tabla 6.7

En este caso se nota un valor de significatividad que es mayor a 0,05. Se procede a quitar la variable situación laboral. Esto se debe también a la obtención de un modelo más parsimonioso.

Se debe aclarar que los pasos anteriores se rehicieron con el modelo corregido que no contempla la situación laboral y pasaron las pruebas.

Con el modelo corregido la nueva tabla de estimación de parámetros es:

Estimaciones de los parámetros

Género de Película Preferido		B	Error típ.	Wald	gl	Sig.	Exp(B)	Intervalo de confianza al 95% para Exp(B)	
								Límite inferior	Límite superior
Drama	Intersección	-,868	,281	9,552	1	,002			
	[catedeedad=1]	1,171	,313	13,963	1	,000	3,225	1,745	5,960
	[catedeedad=2]	1,488	,320	21,618	1	,000	4,428	2,365	8,290
	[catedeedad=3]	,753	,339	4,937	1	,026	2,124	1,093	4,126
	[catedeedad=4]	0 ^b	.	.	0
	[estadciv=0]	-,750	,191	15,375	1	,000	,472	,324	,687
	[estadciv=1]	0 ^b	.	.	0
Comedia	Intersección	,941	,186	25,696	1	,000			
	[catedeedad=1]	-3,606	,400	81,278	1	,000	,027	,012	,059
	[catedeedad=2]	-2,249	,266	71,671	1	,000	,106	,063	,178
	[catedeedad=3]	-1,045	,221	22,292	1	,000	,352	,228	,543
	[catedeedad=4]	0 ^b	.	.	0
	[estadciv=0]	,411	,208	3,922	1	,048	1,508	1,004	2,266
	[estadciv=1]	0 ^b	.	.	0

a. La categoría de referencia es: Acción.

b. Este parámetro se ha establecido a cero porque es redundante.

Tabla 6.8

Como se observa los nuevos valores de significatividad son todos menores a 0,05 y se procede con el análisis. A su vez, los valores positivos de B comparando las distintas categorías de edad con respecto a la categoría 4 (mayores a 60 años) muestran como a menor edad las películas de género dramático se prefieren *relativamente* por sobre las de acción (no es una comparación absoluta). Es decir que aumenta el porcentaje de gente que ve dramas cuando disminuye la edad lo que no quiere decir que más gente vea dramas que películas de acción.

* Luego, se observa el valor de porcentaje global de la Tabla de Clasificación. Se debe comparar este valor con el de porcentaje marginal de la Tabla de Resumen de Procesamientos de Casos. Es imprescindible que el porcentaje global sea mayor que el porcentaje marginal de interés ya que este último corresponde al modelo nulo. Puede que un modelo sea excelente para identificar uno de los valores que puede tomar la variable dependiente y no tanto con otro de los valores.

En nuestro ejemplo:

Clasificación

Observado	Pronosticado			Porcentaje correcto
	Drama	Comedia	Acción	
Drama	82	48	104	35,0%
Comedia	5	275	52	82,8%
Acción	57	128	129	41,1%
Porcentaje global	16,4%	51,3%	32,4%	55,2%

Tabla 6.9

En este caso tanto el valor global (55,2 %) como los demás porcentajes son satisfactorios y superan a los que se obtienen con el modelo nulo que se muestran a continuación en la tabla 6.10:

Resumen del procesamiento de los casos

		N	Porcentaje marginal
Género de Película Preferido	Drama	234	26,6%
	Comedia	332	37,7%
	Acción	314	35,7%
Categoría de Edad	Hasta 30	181	20,6%
	31-45	206	23,4%
	46-60	231	26,3%
	Más de 60	262	29,8%
Estado Civil	Soltero	577	65,6%
	Casado	303	34,4%
Válidos		880	100,0%
Perdidos		0	
Total		880	
Subpoblación		8	

Tabla 6.10

*Por último, se observan los valores de los residuos de Pearson que se ubican en la Tabla de Frecuencias Observadas y Predecidas. Un valor muy alto de Residuo de Pearson denota un patrón de variables independientes que no es bien ajustado por el modelo.

En nuestro ejemplo:

Frecuencias observadas y pronosticadas

Estado Civil	Categoría de Edad	Género de Película Preferido	Frecuencia			Porcentaje	
			Observada	Pronosticada	Residuo de Pearson	Observada	Pronosticada
Soltero	Hasta 30	Drama	32	33	-,135	36,0%	36,6%
		Comedia	6	5	,286	6,7%	6,0%
		Acción	51	51	-,007	57,3%	57,3%
	31-45	Drama	60	59	,146	39,0%	38,4%
		Comedia	29	28	,315	18,8%	17,9%
		Acción	65	67	-,387	42,2%	43,8%
	46-60	Drama	30	29	,285	15,9%	15,1%
		Comedia	89	92	-,501	47,1%	48,9%
		Acción	70	68	,309	37,0%	36,0%
	Más de 60	Drama	4	6	-,716	2,8%	3,9%
		Comedia	112	111	,254	77,2%	76,3%
		Acción	29	29	,078	20,0%	19,7%
Casado	Hasta 30	Drama	52	51	,128	56,5%	55,9%
		Comedia	2	3	-,401	2,2%	2,9%
		Acción	38	38	,007	41,3%	41,3%
	31-45	Drama	30	31	-,249	57,7%	59,4%
		Comedia	3	4	-,739	5,8%	8,6%
		Acción	19	17	,708	36,5%	32,0%
	46-60	Drama	12	13	-,464	28,6%	31,9%
		Comedia	17	14	1,135	40,5%	32,3%
		Acción	13	15	-,656	31,0%	35,8%
	Más de 60	Drama	14	12	,504	12,0%	10,5%
		Comedia	74	75	-,251	63,2%	64,4%
		Acción	29	29	-,079	24,8%	25,1%

Los porcentajes se basan en las frecuencias observadas totales de cada subpoblación.

Tabla 6.11

Se ve que no hay ningún valor de Residuo de Pearson que sea extremadamente alto. Nótese también la verosimilitud entre los porcentajes observados y los pronosticados por el modelo.

*Resulta fundamental aclarar que si bien estos pasos mecánicos que se han mencionado deben ser respetados, se debe utilizar el criterio y el sentido común para obtener un buen modelo. No todos los modelos que pasen estas pruebas son válidos. También es deseable ser criterioso para obtener modelos parsimoniosos que no contengan demasiadas variables explicativas.

6.3.2 Contraste de Hipótesis

En este caso, luego realizados todos los pasos anteriores se procede a verificar si las hipótesis a priori eran correctas o no sucede lo que habíamos supuesto.

En el caso de nuestro modelo:

Estimaciones de los parámetros

Género de Película Preferido		B	Error típ.	Wald	gl	Sig.	Exp(B)	Intervalo de confianza al 95% para Exp(B)	
								Límite inferior	Límite superior
Drama	Intersección	-,868	,281	9,552	1	,002			
	[catedeedad=1]	1,171	,313	13,963	1	,000	3,225	1,745	5,960
	[catedeedad=2]	1,488	,320	21,618	1	,000	4,428	2,365	8,290
	[catedeedad=3]	,753	,339	4,937	1	,026	2,124	1,093	4,126
	[catedeedad=4]	0 ^b	.	.	0
	[estadciv=0]	-,750	,191	15,375	1	,000	,472	,324	,687
	[estadciv=1]	0 ^b	.	.	0
Comedia	Intersección	,941	,186	25,696	1	,000			
	[catedeedad=1]	-3,606	,400	81,278	1	,000	,027	,012	,059
	[catedeedad=2]	-2,249	,266	71,671	1	,000	,106	,063	,178
	[catedeedad=3]	-1,045	,221	22,292	1	,000	,352	,228	,543
	[catedeedad=4]	0 ^b	.	.	0
	[estadciv=0]	,411	,208	3,922	1	,048	1,508	1,004	2,266
	[estadciv=1]	0 ^b	.	.	0

a. La categoría de referencia es: Acción.

b. Este parámetro se ha establecido a cero porque es redundante.

Tabla 6.12

Los valores negativos aumentan a medida que la categoría de edad es menor. Obsérvese que al ocurrir esto disminuirán las chances de que, por ejemplo, una persona joven prefiera las comedias por sobre las películas de acción.

Esto se ve también en la Tabla de Frecuencias observadas y pronosticadas.

Frecuencias observadas y pronosticadas

Estado Civil	Categoría de Edad	Género de Película Preferido	Frecuencia			Porcentaje	
			Observada	Pronosticada	Residuo de Pearson	Observada	Pronosticada
Soltero	Hasta 30	Drama	32	33	-,135	36,0%	36,6%
		Comedia	6	5	,286	6,7%	6,0%
		Acción	51	51	-,007	57,3%	57,3%
	31-45	Drama	60	59	,146	39,0%	38,4%
		Comedia	29	28	,315	18,8%	17,9%
		Acción	65	67	-,387	42,2%	43,8%
	46-60	Drama	30	29	,285	15,9%	15,1%
		Comedia	89	92	-,501	47,1%	48,9%
		Acción	70	68	,309	37,0%	36,0%
Más de 60	Drama	4	6	-,716	2,8%	3,9%	
	Comedia	112	111	,254	77,2%	76,3%	
	Acción	29	29	,078	20,0%	19,7%	
Casado	Hasta 30	Drama	52	51	,128	56,5%	55,9%
		Comedia	2	3	-,401	2,2%	2,9%
		Acción	38	38	,007	41,3%	41,3%
	31-45	Drama	30	31	-,249	57,7%	59,4%
		Comedia	3	4	-,739	5,8%	8,6%
		Acción	19	17	,708	36,5%	32,0%
	46-60	Drama	12	13	-,464	28,6%	31,9%
		Comedia	17	14	1,135	40,5%	32,3%
		Acción	13	15	-,656	31,0%	35,8%
	Más de 60	Drama	14	12	,504	12,0%	10,5%
		Comedia	74	75	-,251	63,2%	64,4%
		Acción	29	29	-,079	24,8%	25,1%

Los porcentajes se basan en las frecuencias observadas totales de cada subpoblación.

Tabla 6.13

Si se comparan la categoría de edad 4 (mayores de 60) con respecto a la categoría de edad 1 (menores a 30) y recordando:

$$\beta_1 = \ln \left(\frac{\text{odds}(X = 1)}{\text{odds}(X = 0)} \right)$$

Fórmula 6.1

en este caso X = 0 sería la categoría 4 y X = 1 la categoría 1.

Entonces operando matemáticamente resulta:

$$e^{-3,6} \times 75/29 \cong 3/38$$

Fórmula 6.2

Este cálculo¹² refleja como las chances de preferir una comedia por sobre una película de acción disminuyen a medida que disminuye la edad (nótese el coeficiente -3,6). Lo que confirma la hipótesis planteada a priori.

¹² El valor aproximado se debe a redondeos de los valores que arroja el modelo.

SECCIÓN II

MODELOS OBTENIDOS A PARTIR DE LA INVESTIGACIÓN

CAPÍTULO 7 - MODELOS OBTENIDOS A PARTIR DE LA INVESTIGACIÓN

Se divide esta sección en dos partes, en una se explican los modelos generales obtenidos analizando la totalidad de la muestra y en la otra se explican los modelos obtenidos que contemplan sólo a la gente más joven. Se procede de esta forma, ya que es razonable pensar que la edad es un factor muy influyente en las elecciones en cuanto al turismo se refieren.

7.1 Modelos Generales referentes a Viajes Turísticos

Estos modelos hacen referencia a las distintas elecciones que se pueden hacer referentes a viajes turísticos. Vale la aclaración que todos cumplen con absolutamente todos los requerimientos estadísticos detallados en las secciones anteriores. Para la elección de las variables independientes principales, se evaluaron más de cincuenta para cada modelo, que surgen de combinaciones de las variables relevadas en el trabajo de campo.

7.1.1 Modelo Referente a la Elección de Tipo de Hospedaje.

La variable Dependiente en este caso es:

- ❖ *¿Utilizaría Hoteles de 3 o más Estrellas en futuros Viajes Turísticos?*

Esta pregunta se responde por *Si* o por *No*.

Para la obtención de las variables independientes se siguieron los pasos explicados en el capítulo 6 y las elegidas resultaron:

- *¿Utilizó Hoteles de 3 o más Estrellas en Viajes Turísticos Pasados?*

Esta pregunta se responde por *Si* o por *No*.

- *¿Es Soltero?*

Esta variable agrupa a todos los casados o en pareja en *no solteros*. De esta manera *Si* corresponde a los solteros y *No* al resto.

7.1.1.1 Planteo de Hipótesis

Experiencias como la de Modelos de Autoscoring relacionados con los seguros automovilísticos nos hacen pensar que los hechos pasados condicionan el futuro.

En este sentido hacemos referencia a que las personas que sufrieron siniestros en el pasado son más propensas a sufrir siniestros en el futuro.

De esta manera, realizando una analogía en nuestro caso, suena razonable esperar que personas que ya han utilizado hoteles de categoría en el pasado, lo vuelvan a hacer en alguna oportunidad futura.

Por lo tanto:

Hipótesis 1:

Las personas que han viajado a hoteles de categoría en el pasado son más propensas a viajar a hoteles de categoría en el futuro.

Con respecto al estado civil de las personas, es lógico pensar que las personas que no están en pareja ni casadas pueden pretender menos comodidades en un viaje.

Por lo tanto:

Hipótesis 2:

Las personas que se encuentran casadas o en pareja son más propensas a viajar a hoteles de categoría en el futuro.

7.1.1.1 Resultados del Modelo

Resumen del procesamiento de los casos

		N	Porcentaje marginal
ViajaráHotel3Estrellas	No	80	42%
	Si	109	58%
Soltero	No	63	33%
	Si	126	67%
ViajóHoteles3Estrellas	No	73	39%
	Si	116	61%
Válidos		189	100%
Perdidos		2	
Total		191	
Subpoblación		4	

Tabla 7.1

En la tabla de arriba se observa que el 58 % de los encuestados planea viajar a hoteles 3 o más estrellas, que el 67 % de los encuestados son solteros y que el 61 % de los encuestados viajaron a hoteles de 3 o más estrellas en el pasado. Se contemplan 189 casos en el análisis.

Con respecto a la clasificación del modelo en cuánto a personas que viajarán a hoteles de categoría (Si) con respecto a los que no lo harán (No):

Clasificación

Observado	Pronosticado		
	No	Si	Porcentaje correcto
No	45	35	56%
Si	15	94	86%
Porcentaje global	32%	68%	74%

Tabla 7.2

Las celdas en diagonal son los casos en que las predicciones son correctas. Se observa que el 56 % de las personas que no viajarán a hoteles de categoría son diagnosticadas correctamente. En el caso de las personas que si lo harán el porcentaje de casos bien diagnosticados es de 86 %.

En el global, el 74 % de los casos son bien diagnosticados. Esto supera en un 16 % al modelo nulo que clasifica todos los casos como viajantes de hoteles de categoría (58 %).

Con respecto a las frecuencias observadas y Pronosticadas:

Frecuencias observadas y pronosticadas

			Frecuencia			Porcentaje	
			Observada	Pronosticada	Residuo de Pearson	Observada	Pronosticada
ViajóHoteles3Estrellas	Soltero	No	6	6	,097	46%	45%
		Si	7	7	-,097	54%	55%
	ViajaráHotel3Estrellas	No	45	45	-,052	75%	75%
		Si	15	15	,052	25%	25%
Si	No	No	6	6	-,075	12%	12%
		Si	44	44	,075	88%	88%
	ViajaráHotel3Estrellas	No	23	23	,045	35%	35%
		Si	43	43	-,045	65%	65%

Los porcentajes se basan en las frecuencias observadas totales de cada subpoblación.

Tabla 7.3

Nótese el buen ajuste del modelo, comparando los valores pronosticados y observados.

Con respecto a las hipótesis planteadas, se observa que las personas que no viajaron a hoteles de categoría en el pasado y son solteras, son las que menos probabilidades tienen de elegir hoteles de categoría en viajes futuros (25 %).

En contraposición, las personas que sí han viajado a hoteles de categoría y están casadas o en pareja, son las que más probabilidades tienen de elegir hoteles de categoría en viajes futuros (88 %).

7.1.1.2 Contraste de Hipótesis

Para realizar un contraste de hipótesis definitivo se analizará la siguiente tabla:

Estimaciones de los parámetros							Intervalo de confianza al 95% para Exp(B)	
ViajaráHotel3Estrellas ^a	B	Error típ.	Wald	gl	Sig.	Exp(B)	Límite inferior	Límite superior
Si	Intersección	,64	,25	6,76	1	,009		
	[Soltero=,00]	1,32	,39	11,31	1	,001	3,75	1,74
	[Soltero=1,00]	0 ^b	.	.	0	.	.	.
	[ViajóHoteles3 Estrellas=,00]	-1,75	,34	25,78	1	,000	,17	,09
	[ViajóHoteles3 Estrellas=1,00]	0 ^b	.	.	0	.	.	.

a. La categoría de referencia es: No.

b. Este parámetro se ha establecido a cero porque es redundante.

Tabla 7.4

De la tabla 7,4, comparando a las personas que no viajaron a hoteles en el pasado contra los que sí lo hicieron, se observa el valor menor a la unidad de e^B ($\text{Exp}(B) = 0,17$), que muestra que es menos probable que una persona que no haya viajado a hoteles en el pasado lo haga en el futuro comparando contra los que sí han ido. Reacomodando los cálculos, una persona que ha viajado a hoteles de categoría en el pasado tiene $1/0,17=5,9$ veces más chances de ir a un hotel de categoría en el futuro que una persona que no ha viajado a hoteles de este tipo. O sea, las chances de elegir hoteles de categoría aumentan 490 % cuando las personas ya han ido a hoteles de este tipo. **Lo que confirma fuertemente la hipótesis 1.**

De la misma manera, de la tabla 7.4, comparando a los no solteros con los solteros, se observa el valor mayor a la unidad de e^B ($\text{Exp}(B) = 3,75$), que muestra que es más probable que las personas “no solteras” opten por un viaje en hoteles de categoría, en comparación con los solteros. Para ser más precisos, una persona “no soltera” presenta 3,75 veces más chances de elegir un hotel de categoría que una soltera. O sea, las chances de elegir hoteles de categoría aumentan 275 % cuando las personas no se encuentran solteras. **Lo que confirma fuertemente la hipótesis 2.**

7.1.1.3 Resumen y Conclusiones del Modelo

Se obtuvo un modelo que trata de predecir si una persona estaría o no dispuesta a utilizar hoteles de categoría en un futuro viaje de placer.

Las cuestiones que resultaron relevantes a la hora de definir esto fueron:

- si la persona en cuestión había utilizado hoteles de este tipo en el pasado o no lo había hecho.
- si la persona en cuestión se encuentra soltera o no.

Se establecieron dos hipótesis:

Hipótesis 1:

Las personas que han viajado a hoteles de categoría en el pasado son más propensas a viajar a hoteles de categoría en el futuro.

Hipótesis 2:

Las personas que se encuentran casadas o en pareja son más propensas a viajar a hoteles de categoría en el futuro.

Se confirmaron ambas hipótesis, ya que se pudo demostrar con significatividad estadística que las personas casadas o en pareja presentan un 275 % más chances de elegir hoteles de lujo en un viaje futuro, comparando con los solteros. A su vez, las personas que ya han utilizado este tipo de alojamiento en el pasado presentan un 490 % más chances de volver a utilizarlo en el futuro, comparando con los que nunca lo hicieron.

Como caso interesante para mencionar del modelo se encuentra el de las personas solteras y que no han viajado a hoteles de categoría, presentando una probabilidad de viajar a hoteles de categoría en un futuro viaje de sólo 25 %. En contraposición, se encuentran las personas casadas o en pareja que sí han viajado a hoteles de categoría en oportunidades anteriores, presentando una probabilidad de volver a utilizar estos hoteles de 88 %.

7.1.2 Modelo Referente a la Elección de Distintos Destinos.

La variable Dependiente en este caso es:

- ❖ *Destino más Importante que viajará excluyendo Europa.*

En este caso se agrupan los destinos de la siguiente manera, ordenados de menor a mayor relevancia:

- Argentina: La *región máxima que visitará* será esta cuando los encuestados hayan respondido que planean viajar a algún destino dentro de la Argentina y no hayan elegido ningún otro destino.
- Países Limítrofes: La *región máxima que visitará* será esta cuando los encuestados hayan respondido querer viajar a los destinos Chile, Bolivia o Perú¹³, Uruguay o Brasil. A su vez, las personas que hayan mencionado cualquiera de los destinos anteriores y también hayan elegido algún destino en Argentina se colocarán en esta categoría pues se priorizan los destinos más “importantes”.
- *Resto del Mundo Sin Europa*¹⁴: La *región máxima que visitará* será esta cuando los encuestados hayan respondido querer viajar a los destinos Colombia, Centroamérica, Norteamérica, Oceanía, África o Asia. A su vez, las personas que hayan mencionado cualquiera de los destinos anteriores y también hayan elegido Argentina o algún país limítrofe se colocarán en esta categoría pues nuevamente se priorizan los destinos más “importantes”.

Para la obtención de las variables independientes se siguieron los pasos explicados en el capítulo 6 y las elegidas resultaron:

- ¿Viajó al Resto de América?

Esta variable se responde por *Si* o por *No*. Se considera “Resto de América” a los destinos Centroamérica y/o Estados Unidos.

- ¿Gasta moderadamente en Viajes Turísticos?

Esta variable se responde por *Si* o por *No*.

7.1.2.1 Planteo de Hipótesis

En este caso, es de esperar que las personas que han viajado a destinos que no son económicos, como Centroamérica o Estados Unidos, sean más propensas a querer volver a esos destinos, o bien visitar lugares como Oceanía o África. La validación del modelo explicado anteriormente fundamenta esto ya que se demostró que las experiencias pasadas condicionan el accionar futuro. A su vez, suponiendo que no hicieran viajes de tanta envergadura, es razonable esperar que al menos viajen a países limítrofes no limitando sus viajes a destinos dentro de la Argentina.

Por lo tanto:

Hipótesis 1:

¹³ Perú no es un país limítrofe de la Argentina pero se lo tomó como tal para simplificar las agrupaciones.

¹⁴ Al destino Europa no se lo considera ya que será evaluado en un modelo particular para este destino.

Las personas que han viajado al Resto de América en el pasado son más propensas a viajar al Resto del Mundo en el futuro y poco propensas a limitarse sólo a viajes locales

En lo que respecta al gasto en turismo, es comprensible esperar que las personas que dicen tener un gasto moderado en el rubro, se encuentren menos ávidas a realizar viajes a destinos muy distantes y exclusivos.

Por ende:

Hipótesis 2:

Las personas que declaran gastar moderadamente en el rubro turismo son menos propensas a viajar al Resto del Mundo y más propensas a realizar viajes a países limítrofes o a destinos locales.

7.1.2.2 Resultados del Modelo

Resumen del procesamiento de los casos

		N	Porcentaje marginal
RegiónMáximaQue	Argentina	53	28%
Viajará	PaísesLimítrofes	64	34%
	Resto Del Mundo SE	73	38%
GastaMedio	No	72	38%
	Si	118	62%
ViajóRestodeAmérica	No	143	75%
	Si	47	25%
Válidos		190	100%
Perdidos		192	
Total		382	
Subpoblación		4	

Tabla 7.5

En la tabla de arriba se observa que el 28 % de los encuestados planea viajar como máximo a la Argentina, el 34 % como máximo a los países limítrofes y el 38 % estaría interesado en viajar al *resto del mundo*, totalizando entre las tres opciones el 100 % de los casos. A su vez, el 62 % de las personas respondieron gastar moderadamente en turismo y el 25 % respondió haber viajado al resto de América. Se contemplan 190 casos.

Con respecto a la clasificación del modelo en cuánto a la elección del máximo destino a realizar se presenta la siguiente tabla:

Clasificación

Observado	Pronosticado			Porcentaje correcto
	Argentina	Países Limítrofes	Resto Del Mundo SE	
Argentina	0	27	26	0%
Países Limítrofes	0	42	22	66%
Resto Del Mundo SE	0	19	54	74%
Porcentaje global	0%	46%	54%	51%

Tabla 7.6

Las celdas en diagonal son los casos en que las predicciones son correctas. Se observa que el 74 % de las personas que viajan al *resto del mundo* son diagnosticadas correctamente, también se observa que el 66 % de las personas interesadas en viajar como máximo a un país limítrofe son bien diagnosticadas. En contraposición, nótese el pésimo porcentaje de aciertos en lo que respecta a los viajantes a la Argentina (0 %). Esto se debe a que en ninguno de los casos la Argentina es la categoría de respuesta predicha con la probabilidad más alta arrojada por el modelo. De querer segmentar a estas personas se tendrían que buscar otras variables predictoras.

En el global, el 51 % de los casos son bien diagnosticados. Esto supera a cualquiera de los modelos nulos que se quisieran plantear con cualquier categoría de referencia.

Con respecto a las frecuencias observadas y pronosticadas:

Frecuencias observadas y pronosticadas

Viajó Resto de América	Gasta Medio	Región Máxima Que Viajará	Frecuencia			Porcentaje	
			Observada	Pronosticada	Residuo de Pearson	Observada	Pronosticada
No	No	Argentina	15	15	-,1	27%	28%
		Países Limítrofes	14	14	,1	25%	25%
		Resto Del Mundo SE	26	26	,0	47%	48%
	Si	Argentina	27	27	,0	31%	30%
		Países Limítrofes	42	42	-,1	48%	48%
		Resto Del Mundo SE	19	19	,0	22%	21%
Si	No	Argentina	3	3	,1	18%	17%
		Países Limítrofes	1	1	-,3	6%	8%
		Resto Del Mundo SE	13	13	,1	76%	76%
	Si	Argentina	8	8	-,1	27%	27%
		Países Limítrofes	7	7	,1	23%	22%
		Resto Del Mundo SE	15	15	-,1	50%	50%

Los porcentajes se basan en las frecuencias observadas totales de cada subpoblación.

Tabla 7.7

Nótese el buen ajuste del modelo, comparando los valores pronosticados y observados.

Con respecto a las hipótesis planteadas, se observa que las personas que no han viajado al resto de América y tienen un gasto moderado en turismo, son las que menos probabilidades tienen de estar interesadas en viajar al *resto del mundo* (22 %).

En contraposición, las personas que si han viajado al resto de América y no son moderados en cuanto al gasto en turismo, son las que más probabilidades tienen de estar interesadas en viajar al *resto del mundo* (76%).

7.1.2.3 Contraste de Hipótesis

Para realizar un contraste de hipótesis definitivo se analizará la siguiente tabla:

Estimaciones de los parámetros

RegiónMáximaQue Viajará ^a		B	Error típ.	Wald	gl	Sig.	Exp(B)	Intervalo de confianza al 95% para Exp(B)	
								Límite inferior	Límite superior
PaísesLimítrofes	Intersección	-,20	,47	,18	1	,67			
	[GastaMedio=,00]	-,56	,42	1,81	1	,18	,57	,25	1,29
	[GastaMedio=1,00]	0 ^b	.	.	0
	[ViajóRestode América=,00]	,66	,51	1,66	1	,20	1,93	,71	5,25
	[ViajóRestode América=1,00]	0 ^b	.	.	0
Resto Del Mundo SE	Intersección	,62	,38	2,63	1	,10	2,45	1,16	5,20
	[GastaMedio=,00]	,90	,38	5,48	1	,02			
	[GastaMedio=1,00]	0 ^b	.	.	0	.			
	[ViajóRestode América=,00]	-,97	,43	5,17	1	,02	,38	,16	,87
	[ViajóRestode América=1,00]	0 ^b	.	.	0	.			

a. La categoría de referencia es: Argentina.

b. Este parámetro se ha establecido a cero porque es redundante.

Tabla 7.8

De la tabla 7.8, comparando a las personas que viajaron al resto de América contra los que no lo hicieron, se observa el valor menor a la unidad de e^B ($\text{Exp}(B) = 0,38$), que muestra que es menos probable que una persona que no ha viajado al resto de América lo haga al *resto del mundo*. Reacomodando los cálculos, una persona que ha viajado al resto de América en el pasado tiene $1/0,38=2,63$ veces más chances de viajar al *resto del mundo* en el futuro. O sea, las chances de

viajar al *resto del mundo* aumentan 163 % cuando las personas han viajado al resto de América. **Todo esto confirma fuertemente la hipótesis 1.**

De la misma manera, de la tabla anterior, comparando a las personas que no declararon ser moderados en sus gastos contra los que sí lo hicieron, se observa el valor mayor a la unidad de e^B ($\text{Exp}(B) = 2,45$), que muestra que es más probable que una persona que no gasta mesuradamente viaje al *resto del mundo*, comparando con el resto de las personas. Para ser más precisos, las personas “no moderadas” presentan 2,45 veces más chances de viajar al *resto del mundo*. O sea, las chances de viajar al *resto del mundo* aumentan un 145 % cuando las personas no declaran realizar gastos moderados en el rubro turismo. A su vez, las diferencias entre los que declaran gastar modestamente o no, no son significativas para la comparación entre los que quieren viajar a los países limítrofes y a la Argentina. Por lo tanto, ya que hay diferencias entre los que viajarán a *resto del mundo* y los que sólo lo harán a la Argentina, y no hay diferencias significativas entre los que viajarán a los países limítrofes y la Argentina, hay diferencias entre los que viajan al *resto del mundo* y los que lo hacen a los países limítrofes por carácter transitivo. **Todo esto confirma fuertemente la hipótesis 2.**

7.1.2.4 Resumen y Conclusiones del Modelo

Se obtuvo un modelo que trata de predecir entre tres tipos de destinos que una persona puede elegir para un viaje turístico. Los destinos se agruparon en tres categorías: Argentina, países limítrofes y *resto del mundo*.

Las cuestiones que resultaron relevantes a la hora de definir el destino a viajar fueron:

- si la persona en cuestión había viajado al resto de América.
- si la persona en cuestión declaraba ser moderado en sus gastos turísticos.

Se establecieron dos hipótesis:

Hipótesis 1:

Las personas que han viajado al Resto de América en el pasado son más propensas a viajar al Resto del Mundo en el futuro y poco propensas a limitarse sólo a viajes locales

Hipótesis 2:

Las personas que declaran gastar moderadamente en el rubro turismo son menos propensas a viajar al Resto del Mundo y más propensas a realizar viajes a países limítrofes o a destinos locales.

Se confirmaron ambas hipótesis, ya que se pudo demostrar con significatividad estadística que las personas que viajaron al resto de América en el pasado,

presentan un 163 % más chances de viajar al *resto del mundo* en un futuro viaje, comparando contra los que no viajaron. A su vez, las personas que no declaran un gasto modesto en turismo presentan un 145 % más chances de viajar al *resto del mundo*, comparando con los que declaran un gasto modesto.

Como caso interesante para mencionar se encuentra el que se muestra en la tabla 7.9, en la que se comparan dos grupos de personas. Uno que no ha viajado al resto de América y declara un gasto moderado y otro que sí ha viajado al resto de América y no es moderado con los gastos. Nótese las importantes diferencias entre los futuros destinos elegidos por ambos:

Viajó Resto de América	Gasto Moderado	Región Máxima Que Viajará	Porcentaje		Cambio
			Observada	Pronosticada	
No	Si	Argentina	31%	30%	
No	Si	Países Límites	48%	48%	
No	Si	Resto Del Mundo SE	22%	21%	
					Cambio
Si	No	Argentina	18%	17%	↓
Si	No	Países Límites	6%	8%	↓
Si	No	Resto Del Mundo SE	76%	76%	↑

Tabla 7.9

7.1.3 Modelo Referente a la Elección del Destino Europa.

La variable Dependiente en este caso es:

- ❖ ¿Planea viajar a Europa?

Esta variable se responde por *Si* o por *No*.

Para la obtención de las variables independientes se siguieron los pasos explicados en el capítulo 6 y las elegidas resultaron:

- ¿Viajó a Chile, Bolivia o Perú?

Esta pregunta se responde por *Si* o por *No*.

- ¿Considera su situación económica muy buena?

Esta variable agrupa a todas las personas que respondieron tener una situación económica mala, regular y buena con la respuesta *No*. Los que respondieron tener una situación económica muy buena se encuentran en la respuesta *Si*.

7.1.3.1 Planteo de Hipótesis

Se puede suponer que las personas que han viajado a destinos como Bolivia o Perú aprecian los viajes culturales más que la media de las personas. No todo el mundo está interesado en estos destinos, y la elección de ellos denota un perfil aventurero y de viajante activo que no se conformaría con descansar en una playa. Por estos motivos, es natural esperar que personas que han viajado a estos destinos, puedan estar más interesados a viajar a un destino con muchos atractivos culturales como Europa, comparando contra los que no lo han hecho.

Por lo tanto:

Hipótesis 1:

Las personas que han viajado a Chile, Bolivia o Perú en el pasado son más propensas a viajar a Europa en el futuro.

Con respecto a la situación económica de las personas, resulta obvio esperar que las personas con una muy buena situación tengan más chances de planear a un viaje a un destino oneroso como puede ser Europa.

Por lo tanto:

Hipótesis 2:

Las personas que declaran tener una situación económica muy buena son más propensas a viajar a Europa en el futuro.

7.1.3.2 Resultados del Modelo

Resumen del procesamiento de los casos

		N	Porcentaje marginal
ViajaráEuropa	No	108	57%
	Si	81	43%
ViajóChileBoliviaOPerú	No	156	83%
	Si	33	17%
EconomíaMuyBuena	No	142	75%
	Si	47	25%
Válidos		189	100%
Perdidos		2	
Total		191	
Subpoblación		4	

Tabla 7.10

En la tabla de arriba se observa que el 43 % de los encuestados tienen pensado viajar a Europa, el 17 % viajó a Chile, Bolivia o Perú y el 25 % de los encuestados respondieron tener una situación económica muy buena. Se contemplan 189 casos en el análisis.

Con respecto a la clasificación del modelo en cuánto a personas que viajarían a Europa (Sí) y a los que no lo harían (No).

Clasificación

Observado	Pronosticado		Porcentaje correcto
	No	Si	
No	84	24	78%
Si	36	45	56%
Porcentaje global	63%	37%	68%

Tabla 7.11

Las celdas en la diagonal de la tabla 7.11 son los casos en que las predicciones son correctas. Se observa que el 78 % de las personas que no planean viajar a Europa son bien diagnosticadas por el modelo. En el caso de los que sí planean hacerlo, el 56 % de los observados son diagnosticados correctamente (no se muestra explícitamente en la tabla, pero es de interés decir que el 65 % de los casos pronosticados como viajantes a Europa lo son en realidad, esto surge del cociente $45/(45+24)$). En el global, el 68 % de los casos son bien diagnosticados. Esto supera claramente al modelo nulo que clasifica todos los casos como viajantes a Europa, con un 43 % de aciertos.

Con respecto a las frecuencias observadas y pronosticadas:

Frecuencias observadas y pronosticadas

				Frecuencia			Porcentaje	
				Observada	Pronosticada	Residuo de Pearson	Observada	Pronosticada
Economía Muy Buena	Viajó Chile Bolivia O Perú	No	84	83	,237	70%	69%	
		Si	36	37	-,237	30%	31%	
	Viajará Europa	No	7	8	-,530	32%	37%	
		Si	15	14	,530	68%	63%	
Economía Buena	Viajó Chile Bolivia O Perú	No	14	15	-,406	39%	42%	
		Si	22	21	,406	61%	58%	
	Viajará Europa	No	3	2	,981	27%	16%	
		Si	8	9	-,981	73%	84%	

Los porcentajes se basan en las frecuencias observadas totales de cada subpoblación.

Tabla 7.12

En este caso, el ajuste del modelo no es excelente como en los casos anteriores, pero estadísticamente significativo y válido para una correcta segmentación.

Con respecto a las hipótesis, se observa que las personas que no viajaron a Chile, Bolivia o Perú y no tienen una muy buena economía son los que menos probabilidad tienen de elegir Europa como destino de un futuro viaje (30 %).

En contraposición, las personas que sí han viajado a Chile, Bolivia o Perú y declaran tener una muy buena situación económica son los que más probabilidades tienen de elegir Europa como destino de un viaje futuro (73 %).

7.1.3.3 Contraste de Hipótesis

Para realizar un contraste de hipótesis definitivo analizaremos la siguiente tabla:

Estimaciones de los parámetros

Viajará Europa ^a	B	Error t.íp.	Wald	gl	Sig.	Exp(B)	Intervalo de confianza al 95% para Exp(B)	
							Límite inferior	Límite superior
No	-1,6	,5	11,7	1	,001			
[Viajó Chile Bolivia O Perú=,00]	1,3	,4	9,7	1	,002	3,7	1,6	8,6
[Viajó Chile Bolivia O Perú=1,00]	,0 ^b	.	.	0
[Economía Muy Buena=,00]	1,1	,4	9,6	1	,002	3,0	1,5	6,2
[Economía Muy Buena=1,00]	,0 ^b	.	.	0

a. La categoría de referencia es: Si.

b. Este parámetro se ha establecido a cero porque es redundante.

Tabla 7.13

De la tabla anterior, comparando a las personas que viajaron a Chile, Bolivia o Perú contra los que no lo hicieron, se observa el valor mayor a la unidad de e^B ($\text{Exp}(B) = 3,7$), que muestra que es más probable que una persona que ha viajado a esos destinos, realice un viaje a Europa. En términos probabilísticos, una persona que ha viajado a esos destinos tiene 3,7 veces más chances de realizar un viaje a Europa, que una persona que no lo ha hecho. O sea, las chances de que una persona planea viajar a Europa aumentan 275 % cuando las personas han viajado a Chile, Bolivia o Perú con anterioridad. **Lo que confirma fuertemente la hipótesis 1.**

De la misma manera, comparando a los que declaran tener una muy buena situación económica con los que no lo hicieron, se observa el valor mayor a la unidad de e^B ($\text{Exp}(B) = 3$), que muestra que es más probable que las personas de economía muy buena planeen viajar a Europa. En términos probabilísticos, una persona de muy buena situación económica tiene 3 veces más chances de planear a viajar a Europa que una persona que no considera tener tan buena situación. O sea, las chances de querer viajar a Europa aumentan un 200 % cuando las personas creen tener una muy buena situación económica. **Lo que confirma fuertemente la hipótesis 2.**

7.1.3.4 Resumen y Conclusiones del modelo

Se obtuvo un modelo que trata de predecir si una persona estaría dispuesta a viajar o no a Europa como destino turístico.

Las cuestiones que resultaron relevantes a la hora de definir esto fueron:

- si la persona en cuestión había viajado a Chile, Bolivia o Perú en el pasado.
- si la persona en cuestión consideraba tener una muy buena situación económica.

Se establecieron dos hipótesis:

Hipótesis 1:

Las personas que han viajado a Chile, Bolivia o Perú en el pasado son más propensas a viajar a Europa en el futuro.

Hipótesis 2:

Las personas que declaran tener una situación económica muy buena son más propensas a viajar a Europa en el futuro.

Se confirmaron ambas hipótesis, ya que se pudo demostrar con significatividad estadística que las personas que viajaron a Chile, Bolivia o Perú presentan un 275 % más chances de elegir viajar a Europa, comparando con los que no lo hicieron. A su vez, las personas que declararon tener una situación económica muy buena

presentan 200 % más chances de elegir Europa como destino, comparando con el resto.

Como caso interesante, se tiene el de las personas que no viajaron a Chile, Bolivia o Perú y que tampoco dicen tener una economía muy buena, con un 30 % de probabilidad de querer viajar a Europa. En oposición, se encuentra el caso de las personas que viajaron a Chile, Bolivia o Perú y también tienen una muy buena situación económica, éstos presentan un 73 % de probabilidades de querer viajar al viejo continente.

7.1.4 Modelo Referente a la Elección de la Modalidad de Organización.

La variable Dependiente en este caso es:

❖ *Modalidad de Organización de Viaje Turístico.*

En este caso se agrupan las modalidades de la siguiente manera, ordenados de menor a mayor independencia:

- Sólo por Agencia de Viajes: En esta categoría se ubican las personas que sólo eligieron la vía más tradicional de organización de un viaje turístico, contactarse con una agencia de viajes, y que esta se ocupe de conseguir y reservar los vuelos (de ser el caso), traslados y el hospedaje.
- Sólo por Internet: En esta categoría se ubican las personas que averiguan o bien reservan tanto vuelos, como traslados u hospedajes a través de la Web.
- Sólo por Cuenta Propia: En esta categoría se encuentran las personas que averiguan lo necesario para sus viajes a través de referencias de amigos, familiares, diarios u otros. En este caso la mayoría de las reservas se realizan por vía telefónica, o bien directamente no se realizan y se emprenden los viajes sin planeamiento previo.

Para la obtención de las variables independientes se siguieron los pasos explicados en el capítulo 6 y las elegidas resultaron:

- *¿Tiene menos de 30 años?*

Esta variable toma el valor *Si* cuando la edad del encuestado se encuentra entre 15 y 30 años. Se agrupan todas las edades mayores a 30 años en *No*. Se aclara que no se relevó a ninguna persona menor a 15 años.

- *¿Organizó viajes por cuenta propia en el pasado?*

En esta variable se agrupan las personas que han contestado que organizaron viajes en el pasado con la opción *Otros*, en *Si*. Se recuerda que los modos de

organizar un viaje de la encuesta eran con agencia de viajes, por Internet y Otros. En “otros”, los encuestados contestaron casi con exclusividad haber organizado todo lo referente a sus viajes, por cuenta propia. Se permitió a su vez que se responda más de una opción, por lo que una persona, por ejemplo, pudo haber contestado que organizó sus viajes por Internet y también con “otros”, quedando agrupado bajo la respuesta *Si*. Los que no seleccionaron la opción otros, fueron agrupados en *No*.

7.1.4.1 Planteo de Hipótesis

En cuanto a la edad de los encuestados, es natural pensar que a mayor edad, menores son las posibilidades de elegir organizaciones de viajes ya sea por Internet, o por cuenta propia. Es más lógico pensar que las personas más grandes, se inclinarán más por la tradicional y más cómoda opción de que una agencia de turismo organice todo lo relevante a sus viajes.

Por lo tanto:

Hipótesis 1:

Las personas de más de 30 años son más propensas a organizar sus viajes mediante agencias de viaje, y menos propensas a hacerlo por Internet o por cuenta propia.

A su vez, es de esperar que las personas que ya han organizado viajes por cuenta propia en oportunidades pasadas, tengan más chances de repetir esto que los demás, o bien optar por una opción intermedia que puede ser Internet. No parece muy probable que su opción predilecta sea la de organizar futuros viajes solamente por agencias de viaje.

Por lo tanto:

Hipótesis 2:

Las personas que han organizado sus viajes por cuenta propia en el pasado, son más propensas a volver a hacerlo, o bien optar por una opción intermedia como puede ser Internet. Este tipo de personas será poco propenso a utilizar solamente agencias de viaje como modalidad de organización.

7.1.4.2 Resultados del Modelo

Resumen del procesamiento de los casos

		N	Porcentaje marginal
OrganizaráSusViajes	Sólo Por Agencia de Viajes	50	38%
	Sólo Por Internet	57	44%
	Sólo Por Cuenta Propia	23	18%
EdadMenorA30	No	45	35%
	Si	85	65%
OrganizóPorCuenta Propia	No	90	69%
	Si	40	31%
Válidos		130	100%
Perdidos		1	
Total		131	
Subpoblación		4	

Tabla 7.14

En la tabla de arriba se observa que el 38 % de los encuestados sólo organizarán sus viajes por Agencia, el 44 % sólo lo hará por Internet y el 18 % restante sólo por cuenta propia, totalizando el 100 % de los casos. Además, se observa que el 65 % de los relevados tienen menos de treinta años y el 31 % ya organizó viajes por su propia cuenta. Se contemplaron 130 casos.

Con respecto a la clasificación del modelo en cuanto a la elección de la modalidad de organización se presenta la siguiente tabla:

Clasificación

Observado	Pronosticado			
	Sólo Por Agencia de Viajes	Sólo Por Internet	Sólo Por Cuenta Propia	Porcentaje correcto
Sólo Por Agencia de Viajes	28	17	5	56%
Sólo Por Internet	12	40	5	70%
Sólo Por Cuenta Propia	5	5	13	57%
Porcentaje global	35%	48%	18%	62%

Tabla 7.15

Las celdas en diagonal son los casos en que las predicciones son correctas. Se observa que el 56 % de las personas que eligen las agencias de viajes son diagnosticadas correctamente, el 70 % de las personas que optan por Internet son bien diagnosticadas y el 57 % de los “cuenta-propistas” se pronostican adecuadamente. En el global, el 62 % de los casos son bien diagnosticados. Esto supera abrumadoramente a cualquier modelo nulo que se quiera plantear.

Con respecto a las frecuencias observadas y pronosticadas:

Frecuencias observadas y pronosticadas

EdadMenorA30	OrganizóPor CuentaPropia	OrganizaráSusViajes	Frecuencia			Porcentaje	
			Observada	Pronosticada	Residuo de Pearson	Observada	Pronosticada
No	No	Sólo Por Agencia de Viajes	18	18	,11	64%	63%
		Sólo Por Internet	10	10	,14	36%	34%
		Sólo Por Cuenta Propia	0	1	-,82	0%	2%
	Si	Sólo Por Agencia de Viajes	10	10	-,14	59%	61%
		Sólo Por Internet	2	2	-,25	12%	14%
		Sólo Por Cuenta Propia	5	4	,36	29%	26%
Si	No	Sólo Por Agencia de Viajes	17	17	-,08	27%	28%
		Sólo Por Internet	40	40	-,10	65%	65%
		Sólo Por Cuenta Propia	5	4	,32	8%	7%
	Si	Sólo Por Agencia de Viajes	5	5	,15	22%	20%
		Sólo Por Internet	5	5	,19	22%	20%
		Sólo Por Cuenta Propia	13	14	-,28	57%	59%

Los porcentajes se basan en las frecuencias observadas totales de cada subpoblación.

Tabla 7.16

Nótese el ajuste satisfactorio del modelo comparando los porcentajes observados y pronosticados.

Con respecto a las hipótesis, nótese que las personas que no son menores a treinta años y no han organizado viajes por cuenta propia, presentan altas probabilidades pronosticadas de contactar una agencia (63 %), menos probabilidades de utilizar Internet (34 %), y escasas probabilidades pronosticadas de organizar viajes por cuenta propia (2%).

En oposición, se encuentra el caso de las personas de menos de treinta años que han organizado viajes por cuenta propia, con moderadas probabilidades de contactar una agencia de viajes (20 %), también modestas probabilidades de utilizar Internet (20%) y altas probabilidades de repetir y utilizar su propia gestión (59%).

7.1.4.3 Contraste de Hipótesis

Para realizar un contraste de hipótesis definitivo se analizará la tabla 7.17:

Estimaciones de los parámetros

OrganizaráSusViajes ^a	B	Error típ.	Wald	gl	Sig.	Exp(B)	Intervalo de confianza al 95% para Exp(B)	
							Límite inferior	Límite superior
Sólo Por Agencia de Viajes	Intersección	-1,1	,5	4,9	1	,03		
	[EdadMenorA30=,00]	1,9	,7	8,7	1	,00	6,9	1,9
	[EdadMenorA30=1,00]	,0 ^b	.	.	0	.	.	.
	[OrganizóPorCuenta Propia=,00]	2,4	,6	14,4	1	,00	11,5	3,3
	[OrganizóPorCuenta Propia=1,00]	,0 ^b	.	.	0	.	.	.
Sólo Por Internet	Intersección	-1,1	,5	4,8	1	,03		
	[EdadMenorA30=,00]	,5	,7	,5	1	,50	1,6	,4
	[EdadMenorA30=1,00]	,0 ^b	.	.	0	.	.	.
	[OrganizóPorCuenta Propia=,00]	3,3	,7	25,4	1	,00	27,3	7,5
	[OrganizóPorCuenta Propia=1,00]	,0 ^b	.	.	0	.	.	.

a. La categoría de referencia es: Sólo Por Cuenta Propia.

b. Este parámetro se ha establecido a cero porque es redundante.

Tabla 7.17

De la tabla anterior, comparando a las personas no menores a treinta años con el resto, se observa el valor mayor a la unidad de e^B ($\text{Exp}(B) = 6,9$), que muestra que es mucho más probable que una persona mayor a 30 años se incline por contactar una agencia de viajes por sobre gestionar sus viajes personalmente. En términos estadísticos, una persona mayor a 30 años tiene 6,9 veces más chances de utilizar una agencia de viajes, comparando con una persona de menos de 30 años. O sea, las chances de elegir una agencia de viajes por sobre la organización por cuenta propia aumentan 590 % cuando las personas tienen más de 30 años. Nótese que no se puede demostrar que las chances aumenten significativamente ($\text{Sig.} = 0,5$), cuando se comparan las chances de elegir Internet por sobre organización por cuenta propia. Pero, por carácter transitivo, las chances de elegir una agencia de viajes por sobre Internet, aumentan también significativamente cuando se es mayor a 30 años. **Lo que confirma fuertemente la hipótesis 1.**

De la misma manera, de la tabla anterior, comparando a los que han organizado por cuenta propia con los que no lo han hecho, se observa el valor mayor a la unidad de e^B ($\text{Exp}(B) = 11,5$), que muestra que es más probable que los que no son “cuenta-propistas” elijan utilizar una agencia de viajes por sobre gestionar sus viajes por si mismos. En términos estadísticos, se tienen 11,5 veces más chances de elegir una agencia de viajes por sobre propia organización, cuando no se ha organizado por cuenta propia en el pasado. O sea, las chances de optar por una agencia de viajes aumentan un 1050 % cuando las personas no han optado por organizar viajes pasados por cuenta propia.

Algo similar, pero en mayor proporción, sucede cuando se compara Internet con “cuenta propia”. En este caso, las chances de optar por organización vía Internet aumentan un 2630 % cuando las personas no han optado por organizar viajes pasados por cuenta propia.

En este caso sólo se puede confirmar parte de la hipótesis y se rechaza la parte restante. Es cierto que las personas que organizaron viajes por cuenta propia son más propensas a volver a hacerlo en el futuro, pero no es cierto que se inclinarán más a la organización por Internet ni que serán poco propensas a utilizar agencias de viajes. En este sentido, la edad resulta un factor más importante.

7.1.4.4 Resumen y Conclusiones del Modelo

Se obtuvo un modelo que trata de predecir entre tres tipos distintos de modalidad de organización de viajes.

Las cuestiones que resultaron relevantes a la hora de definir esto fueron:

- si la persona en cuestión tenía más de treinta años o no lo era.
- si la persona en cuestión había organizado viajes por cuenta propia o no lo había hecho.

Se establecieron dos hipótesis:

Hipótesis 1:

Las personas de más de 30 años son más propensas a organizar sus viajes mediante agencias de viaje, y menos propensas a hacerlo por Internet o por cuenta propia.

Hipótesis 2:

Las personas que han organizado sus viajes por cuenta propia en el pasado, son más propensas a volver a hacerlo, o bien optar por una opción intermedia como puede ser Internet. Este tipo de personas será poco propenso a utilizar solamente agencias de viaje como modalidad de organización.

Se pudo confirmar la primera hipótesis, ya que se pudo demostrar que, a la hora de organizar un viaje, las chances de elegir una agencia por sobre la organización por cuenta propia aumentan 590 % cuando las personas tienen más de 30 años

La segunda hipótesis fue sólo confirmada parcialmente, ya que se pudo probar que las chances de elegir la modalidad “cuenta-propista” aumentan considerablemente cuando ya se ha organizado por esta vía en el pasado. Sin embargo, no es cierto mayoritariamente, que las personas se inclinarán más a la organización por Internet ni que serán poco propensas a utilizar agencias de viajes.

Como caso interesante para mencionar del modelo, se encuentra el de las personas que no son menores a treinta años y no han organizado viajes por cuenta propia, que presentan altas probabilidades pronosticadas de contactar una

agencia (63 %), menos probabilidades de utilizar Internet (34 %), y escasas probabilidades pronosticadas de organizar viajes por cuenta propia (2%).

En oposición, se encuentra el caso de las personas de menos de treinta años que han organizado viajes por cuenta propia, con moderadas probabilidades de contactar una agencia de viajes (20 %), también modestas probabilidades de utilizar Internet (20%) y altas probabilidades de repetir y utilizar su propia gestión (59%).

7.2 Modelos para Jóvenes referentes a Viajes Turísticos

7.2.1 Modelo Referente a la Elección de Tipo de Hospedaje.

La variable Dependiente en este caso es:

- ❖ *¿Utilizaría Hoteles de 3 o más Estrellas en futuros Viajes Turísticos?*

Esta pregunta se responde por *Si* o por *No*.

Para la obtención de las variables independientes se siguieron los pasos explicados en el capítulo 6 y las elegidas resultaron:

- *¿Utilizó Hoteles de 3 o más Estrellas en Viajes Turísticos Pasados?*

Esta pregunta se responde por *Si* o por *No*.

- *¿Viajó a Chile, Bolivia o Perú?*

Esta pregunta se responde por *Si* o por *No*.

7.2.1.1 Planteo de Hipótesis

Como se ha explicado en el modelo general, experiencias como la de Modelos de Autoscoreing relacionados con los seguros automovilísticos nos hacen pensar que los hechos pasados condicionan el futuro. En este sentido hacemos referencia a que las personas que sufrieron siniestros en el pasado son más propensas a sufrir siniestros en el futuro.

De esta manera, realizando una analogía en nuestro caso, suena razonable esperar que personas que ya han utilizado hoteles de categoría en el pasado, lo vuelvan a hacer en alguna oportunidad futura.

Por lo tanto:

Hipótesis 1:

Las personas que han viajado a hoteles de categoría en el pasado son más propensas a viajar a hoteles de categoría en el futuro.

Como se ha explicado en los modelos generales, se puede suponer que las personas que han viajado a destinos como Bolivia o Perú tengan un perfil aventurero y de viajante activo que no presta demasiada atención en las comodidades. Por estos motivos, es natural esperar que personas que han viajado a estos destinos, puedan estar menos interesados en viajar a hoteles de lujo, comparando contra los que no lo han hecho.

Por lo tanto:

Hipótesis 2:

Las personas que han viajado a Chile, Bolivia o Perú son menos propensas a viajar a hoteles de categoría en el futuro.

7.2.1.2 Resultados del Modelo

Resumen del procesamiento de los casos

		N	Porcentaje marginal
ViajaráHotel3Estrellas	No	68	53%
	Si	60	47%
ViajóChileBoliviaOPerú	No	106	83%
	Si	22	17%
ViajóHotel3Estrellas	No	59	46%
	Si	69	54%
Válidos		128	100%
Perdidos		0	
Total		128	
Subpoblación		4	

Tabla 7.18

En la tabla de arriba se observa que el 47 % de los encuestados planea viajar a hoteles 3 o más estrellas (11 % menos que en el modelo general, que involucraba también a los mayores de 30 años), que el 17 % de los encuestados viajaron a Chile, Bolivia o Perú y que el 54 % de los encuestados viajaron a hoteles de 3 o más estrellas en el pasado. Se contemplan 128 casos en el análisis.

Con respecto a la clasificación del modelo en cuánto a personas que viajarán a hoteles de categoría (Si) con respecto a los que no lo harán (No):

Clasificación

Observado	Pronosticado		Porcentaje correcto
	No	Si	
No	51	17	75%
Si	19	41	68%
Porcentaje global	55%	45%	72%

Tabla 7.19

Las celdas en diagonal son los casos en que las predicciones son correctas. Se observa que el 75 % de las personas que no viajarán a hoteles de categoría son diagnosticadas correctamente. En el caso de las personas que si lo harán el porcentaje de casos bien diagnosticados es de 68 %.

En el global, el 72 % de los casos son bien diagnosticados. Esto supera en un 25 % al modelo nulo que clasifica todos los casos como viajeros de hoteles de categoría (47 %).

Con respecto a las frecuencias observadas y Pronosticadas:

Frecuencias observadas y pronosticadas

			Frecuencia			Porcentaje	
			Observada	Pronosticada	Residuo de Pearson	Observada	Pronosticada
ViajóHotel3Estrellas	No	ViajóChileBoliviaOPerú	34	34	,005	71%	71%
		ViajaráHotel3Estrellas	14	14	-,005	29%	29%
	Si	No	10	10	-,015	91%	91%
		Si	1	1	,015	9%	9%
Si	No	No	17	17	-,004	29%	29%
		Si	41	41	,004	71%	71%
	Si	No	7	7	,009	64%	64%
		Si	4	4	-,009	36%	36%

Los porcentajes se basan en las frecuencias observadas totales de cada subpoblación.

Tabla 7.20

Nótese el buen ajuste del modelo, o los valores pronosticados y observados.

Con respecto a las hipótesis planteadas, se observa que las personas que no viajaron a hoteles de categoría en el pasado y si viajaron a Chile, Bolivia o Perú, son las que menos probabilidades tienen de elegir hoteles de categoría en viajes futuros (9 %).

En contraposición, las personas que sí han viajado a hoteles de categoría y no viajaron a Chile, Bolivia o Perú, son las que más probabilidades tienen de elegir hoteles de categoría en viajes futuros (71 %).

7.2.1.3 Contraste de Hipótesis

Para realizar un contraste de hipótesis definitivo se analizará la siguiente tabla:

Estimaciones de los parámetros

ViajaráHotel3Estrellas ^a	B	Error típ.	Wald	gl	Sig.	Exp(B)	Intervalo de confianza al 95% para Exp(B)	
							Límite inferior	Límite superior
Si								
Intersección	-,55	,55	1,02	1	,31			
[ViajóChileBolivia OPerú=,00]	1,43	,58	6,01	1	,01	4,19	1,33	13,18
[ViajóChileBolivia OPerú=1,00]	0 ^b	.	.	0
[ViajóHotel3Estrellas=,00]	-1,76	,40	19,00	1	,00	,17	,08	,38
[ViajóHotel3 Estrellas=1,00]	0 ^b	.	.	0

a. La categoría de referencia es: No.

b. Este parámetro se ha establecido a cero porque es redundante.

Tabla 7.21

De la tabla anterior, comparando a las personas que no viajaron a hoteles en el pasado contra los que sí lo hicieron, se observa el valor menor a la unidad de e^B

($\text{Exp}(B) = 0,17$), que muestra que es menos probable que una persona que no haya viajado a hoteles en el pasado, lo haga en el futuro, comparando contra los que sí han ido. Reacomodando los cálculos, una persona que ha viajado a hoteles de categoría en el pasado tiene $1/0,17=5,9$ veces más chances de ir a un hotel de categoría en el futuro, que una persona que no ha viajado a hoteles de este tipo. O sea, las chances de elegir hoteles de categoría aumentan 490 % cuando las personas ya han ido a hoteles de este tipo. **Lo que confirma fuertemente la hipótesis 1.**

De la misma manera, de la tabla anterior, comparando a los que viajaron a Chile, Bolivia o Perú contra los que no, se observa el valor mayor a la unidad de e^B ($\text{Exp}(B) = 4,19$), que muestra que es más probable que las personas que no viajaron a esos destinos opten por un viaje en hoteles de categoría, en comparación con los que sí lo hicieron. Para ser más precisos, una persona que no viajó a esos destinos presenta 4,19 veces más chances de elegir un hotel de categoría que la que sí viajó. O sea, las chances de elegir hoteles de categoría aumentan 319 % cuando las personas no viajaron a Chile, Bolivia o Perú. **Lo que confirma fuertemente la hipótesis 2.**

7.2.1.4 Resumen y Conclusiones del Modelo

Se obtuvo un modelo que trata de predecir si una persona estaría o no dispuesta a utilizar hoteles de categoría en un futuro viaje de placer.

Las cuestiones que resultaron relevantes a la hora de definir esto fueron:

- si la persona en cuestión había utilizado hoteles de este tipo en el pasado o no lo había hecho.
- si la persona en cuestión había viajado a Chile, Bolivia o Perú o no lo había hecho.

Se establecieron dos hipótesis:

Hipótesis 1:

Las personas que han viajado a hoteles de categoría en el pasado son más propensas a viajar a hoteles de categoría en el futuro.

Hipótesis 2:

Las personas que han viajado a Chile, Bolivia o Perú son menos propensas a viajar a hoteles de categoría en el futuro.

Se confirmaron ambas hipótesis, ya que se pudo demostrar con significatividad estadística que las personas que no viajaron a Chile, Bolivia o Perú presentan un 319 % más chances de elegir hoteles de lujo en un viaje futuro, comparando con los que sí lo hicieron. A su vez, las personas que ya han utilizado este tipo de alojamiento en el pasado presentan un 490 % más chances de volver a utilizarlo en el futuro, comparando con los que nunca lo hicieron.

Como caso interesante para mencionar del modelo se encuentra el de las personas que han viajado a Chile, Bolivia o Perú y que no han viajado a hoteles de categoría, presentando una probabilidad de viajar a hoteles de categoría en un futuro viaje de sólo 9 %. En contraposición, se encuentran las personas que sí han viajado a esos destinos y que sí han viajado a hoteles de categoría en oportunidades anteriores, presentando una probabilidad de volver a utilizar estos hoteles de 71 %.

7.2.2 Modelo Referente a la Elección de Distintos Destinos.

La variable Dependiente en este caso es:

- ❖ *Destino más Importante que viajará excluyendo Europa.*

En este caso se agrupan los destinos de la siguiente manera, ordenados de menor a mayor relevancia:

- Argentina: La *región máxima que visitará* será esta cuando los encuestados hayan respondido que planean viajar a

algún destino dentro de la Argentina y no hayan elegido ningún otro destino.

- Países Limítrofes: La *región máxima que visitará* será esta cuando los encuestados hayan respondido querer viajar a los destinos Chile, Bolivia o Perú¹⁵, Uruguay o Brasil. A su vez, las personas que hayan mencionado cualquiera de los destinos anteriores y también hayan elegido algún destino en Argentina se colocarán en esta categoría pues se priorizan los destinos más “importantes”.
- *Resto del Mundo Sin Europa*¹⁶: La *región máxima que visitará* será esta cuando los encuestados hayan respondido querer viajar a los destinos Colombia, Centroamérica, Norteamérica, Oceanía, África o Asia. A su vez, las personas que hayan mencionado cualquiera de los destinos anteriores y también hayan elegido Argentina o algún país limítrofe se colocarán en esta categoría pues nuevamente se priorizan los destinos más “importantes”.

Para la obtención de las variables independientes se siguieron los pasos explicados en el capítulo 6 y la elegida resultó:

- *¿Viajó al Resto de América?*

Esta variable se responde por *Si* o por *No*. Se considera “Resto de América” a los destinos Centroamérica y/o Estados Unidos.

7.2.2.1 Planteo de Hipótesis

En este caso, es de esperar que las personas que han viajado a destinos que no son económicos, como Centroamérica o Estados Unidos, sean más propensas a querer volver a esos destinos, o bien visitar lugares como Oceanía o África. La validación del modelo explicado anteriormente fundamenta esto ya que se demostró que las experiencias pasadas condicionan el accionar futuro. A su vez, suponiendo que no hicieran viajes de tanta envergadura, es razonable esperar que al menos viajen a países limítrofes no limitando sus viajes a destinos dentro de la Argentina.

Por lo tanto:

Hipótesis:

Las personas que han viajado al Resto de América en el pasado son más propensas a viajar al Resto del Mundo en el futuro.

¹⁵ Perú no es un país limítrofe de la Argentina pero se lo tomó como tal para simplificar las agrupaciones.

¹⁶ Al destino Europa no se lo considera ya que será evaluado en un modelo particular para este destino.

7.2.2.2 Resultados del Modelo

Resumen del procesamiento de los casos

		N	Porcentaje marginal
RegiónMáximaQue	Argentina	29	23%
Viajará	Países Limítrofes	47	37%
	Resto del Mundo SE	52	41%
ViajóRestodeAmérica	No	100	78%
	Si	28	22%
Válidos		128	100%
Perdidos		0	
Total		128	
Subpoblación		2	

Tabla 7.22

En la tabla de arriba se observa que el 23 % de los encuestados planea viajar como máximo a la Argentina, el 37 % como máximo a los países limítrofes y el 41 % estaría interesado en viajar al *resto del mundo*, totalizando entre las tres opciones el 100 % de los casos. A su vez, el 22 % respondió haber viajado al resto de América. Se contemplan 128 casos.

Con respecto a la clasificación del modelo en cuánto a la elección del máximo destino a realizar se presenta la siguiente tabla:

Clasificación

Observado	Pronosticado			Porcentaje correcto
	Argentina	Países Limítrofes	Resto del Mundo SE	
Argentina	0	22	7	0%
Países Limítrofes	0	45	2	96%
Resto del Mundo SE	0	33	19	37%
Porcentaje global	0%	78%	22%	50%

Tabla 7.23

Las celdas en diagonal son los casos en que las predicciones son correctas. Se observa que sólo el 37 % de las personas que viajan al *resto del mundo* son diagnosticadas correctamente, también se observa que el 96 % de las personas interesadas en viajar como máximo a un país limítrofe son bien diagnosticadas. En contraposición, nótese el pésimo porcentaje de aciertos en lo que respecta a los viajantes a la Argentina (0 %). Esto se debe a que en ninguno de los casos la Argentina es la categoría de respuesta predicha con la probabilidad más alta arrojada por el modelo. De querer segmentar a estas personas se tendrían que buscar otras variables predictoras.

En el global, el 50 % de los casos son bien diagnosticados. Esto supera a cualquiera de los modelos nulos que se quisieran plantear con cualquier categoría de referencia.

Con respecto a las frecuencias observadas y pronosticadas:

Frecuencias observadas y pronosticadas

ViajóResto de América	RegiónMáximaQue Viajará	Frecuencia			Porcentaje	
		Observada	Pronosticada	Residuo de Pearson	Observada	Pronosticada
No	Argentina	22	22	,000	22%	22%
	Países Limítrofes	45	45	,000	45%	45%
	Resto del Mundo SE	33	33	,000	33%	33%
Si	Argentina	7	7	,000	25%	25%
	Países Limítrofes	2	2	,000	7%	7%
	Resto del Mundo SE	19	19	,000	68%	68%

Los porcentajes se basan en las frecuencias observadas totales de cada subpoblación.

Tabla 7.24

Nótese el buen ajuste del modelo, comparando los valores pronosticados y observados.

Con respecto a las hipótesis planteadas, se observa que las personas que no han viajado al resto de América, tienen menos probabilidades de estar interesadas en viajar al *resto del mundo* (33% versus 68%).

En cuanto a las personas que no han viajado al resto de América, la opción de viajes a países limítrofes es la que se observa con mayor probabilidad (45% versus 7%)

7.2.2.3 Contraste de Hipótesis

Para realizar un contraste de hipótesis definitivo se analizará la siguiente tabla:

Estimaciones de los parámetros

Región Máxima Que Viajará ^a		B	Error típ.	Wald	gl	Sig.	Exp(B)	Intervalo de confianza al 95% para Exp(B)	
								Límite inferior	Límite superior
Argentina	Intersección	-1,00	,44	5,1	1	,024			
	[Viajó Resto de América=,00]	,59	,52	1,3	1	,255	1,8	,7	5,0
	[Viajó Resto de América=1,00]	,00 ^b	.	.	0
Países Limítrofes	Intersección	-2,25	,74	9,2	1	,002	13,0		
	[Viajó Resto de América=,00]	2,56	,78	10,8	1	,001		2,8	59,5
	[Viajó Resto de América=1,00]	,00 ^b	.	.	0

a. La categoría de referencia es: Resto del Mundo SE.

b. Este parámetro se ha establecido a cero porque es redundante.

Tabla 7.25

De la tabla anterior, comparando a las personas que viajaron al resto de América contra los que no lo hicieron, se observa el valor mayor a la unidad de e^B ($Exp(B) = 13$), que muestra que es más probable que una persona que no ha viajado al resto de América, elija los países limítrofes por sobre el *resto del mundo*. En términos estadísticos, una persona que ha viajado al resto de América en el pasado, tiene 13 veces más chances de querer viajar al *resto del mundo*, comparando contra los que no han viajado. O sea, las chances de querer viajar al *resto del mundo* aumentan 1200 % cuando las personas han viajado al resto de América. **Todo esto confirma fuertemente la hipótesis 1.**

7.2.2.4 Resumen y Conclusiones del Modelo

Se obtuvo un modelo que trata de predecir entre tres tipos de destinos que una persona puede elegir para un viaje turístico. Los destinos se agruparon en tres categorías: Argentina, países limítrofes y *resto del mundo*.

Las cuestiones que resultaron relevantes a la hora de definir el destino a viajar fueron:

- si la persona en cuestión había viajado al resto de América.

Se establecieron la hipótesis:

Hipótesis:

Las personas que han viajado al Resto de América en el pasado son más propensas a viajar al Resto del Mundo en el futuro.

Se pudo confirmar la hipótesis, ya que se pudo demostrar con significatividad estadística que las personas que viajaron al resto de América en el pasado, presentan un 1200 % más chances de viajar al *resto del mundo* en un futuro viaje, comparando contra los que no viajaron, si comparamos con limitarse a los países limítrofes.

Como caso interesante para mencionar se encuentra el que se muestra en la siguiente tabla, en la que se comparan dos grupos de personas. Uno que no ha viajado al resto de América y otro que sí lo ha hecho. Nótese las importantes diferencias entre los futuros destinos elegidos por ambos:

Viajó Resto de América	Región Máxima Que Viajará	Porcentaje	
		Observada	Pronosticada
No	Argentina	22%	22%
No	PaísesLimítrofes	45%	45%
No	Resto Del Mundo SE	33%	33%
Si	Argentina	25%	25%
Si	PaísesLimítrofes	7%	7%
Si	Resto Del Mundo SE	68%	68%

Tabla 7.26

7.2.3 Modelo Referente a la Elección del Destino Europa.

La variable Dependiente en este caso es:

❖ *¿Planea viajar a Europa?*

Esta variable se responde por *Si* o por *No*.

Para la obtención de las variables independientes se siguieron los pasos explicados en el capítulo 6 y las elegidas resultaron:

- *¿Viajó a Chile, Bolivia o Perú?*

Esta pregunta se responde por *Si* o por *No*.

- ¿Considera su situación económica muy buena?

Esta variable agrupa a todas las personas que respondieron tener una situación económica mala, regular y buena con la respuesta *No*. Los que respondieron tener una situación económica muy buena se encuentran en la respuesta *Si*.

7.2.3.1 Planteo de Hipótesis

Como se indicó en el modelo general, se puede suponer que las personas que han viajado a destinos como Bolivia o Perú aprecian los viajes culturales más que la media de las personas. No todo el mundo está interesado en estos destinos, y la elección de ellos denota un perfil aventurero y de viajante activo que no se conformaría con descansar en una playa. Por estos motivos, es natural esperar que personas que han viajado a estos destinos, puedan estar más interesados a viajar a un destino con muchos atractivos culturales como Europa, comparando contra los que no lo han hecho.

Por lo tanto:

Hipótesis 1:

Las personas que han viajado a Chile, Bolivia o Perú en el pasado son más propensas a viajar a Europa en el futuro.

Con respecto a la situación económica de las personas, cómo también ha sido indicado en el modelo general, resulta obvio esperar que las personas con una muy buena situación tengan más chances de planear a un viaje a un destino oneroso como puede ser Europa.

Por lo tanto:

Hipótesis 2:

Las personas que declaran tener una situación económica muy buena son más propensas a viajar a Europa en el futuro.

7.2.3.2 Resultados del Modelo

Resumen del procesamiento de los casos

		N	Porcentaje marginal
ViajaráEuropa	No	71	55%
	Si	57	45%
ViajóChileBoliviaOPerú	No	106	83%
	Si	22	17%
EconomíaMuyBuena	No	100	78%
	Si	28	22%
Válidos		128	100%
Perdidos		0	
Total		128	
Subpoblación		4	

Tabla 7.27

En la tabla de arriba se observa que el 45 % de los encuestados tienen pensado viajar a Europa, el 17 % viajó a Chile, Bolivia o Perú y el 22 % de los encuestados respondieron tener una situación económica muy buena. Se contemplan 128 casos en el análisis.

Con respecto a la clasificación del modelo en cuánto a personas que viajarían a Europa (Sí) y a los que no lo harían (No).

Clasificación

Observado	Pronosticado		Porcentaje correcto
	No	Si	
No	56	15	79%
Si	29	28	49%
Porcentaje global	66%	34%	66%

Tabla 7.28

Las celdas en la diagonal son los casos en que las predicciones son correctas. Se observa que el 79 % de las personas que no planean viajar a Europa son bien diagnosticadas por el modelo. En el caso de los que sí planean hacerlo, el 49 % de los casos observados son diagnosticados correctamente (no se muestra explícitamente en la tabla, pero es de interés decir que el 65 % de los casos pronosticados como viajantes a Europa lo son en realidad, esto surge del cociente $28/(28+15)$). En el global, el 66 % de los casos son bien diagnosticados. Esto supera claramente al modelo nulo que clasifica todos los casos como viajantes a Europa, con un 45 % de aciertos.

Con respecto a las frecuencias observadas y pronosticadas:

Frecuencias observadas y pronosticadas

Economía Muy Buena	Viajó Chile Bolivia O Perú	Viajará Europa	Frecuencia			Porcentaje	
			Observada	Pronosticada	Residuo de Pearson	Observada	Pronosticada
No	No	No	56	56	-,021	66%	66%
		Si	29	29	,021	34%	34%
	Si	No	5	5	,050	33%	33%
		Si	10	10	-,050	67%	67%
Si	No	No	9	9	,040	43%	42%
		Si	12	12	-,040	57%	58%
	Si	No	1	1	-,095	14%	16%
		Si	6	6	,095	86%	84%

Los porcentajes se basan en las frecuencias observadas totales de cada subpoblación.

Tabla 7.29

En este caso, el ajuste es mucho mejor que en el modelo general, estadísticamente significativo y válido para una correcta segmentación.

Con respecto a las hipótesis, se observa que las personas que no viajaron a Chile, Bolivia o Perú y no tienen una muy buena economía son los que menos probabilidad tienen de elegir Europa como destino de un futuro viaje (34 %).

En contraposición, las personas que sí han viajado a Chile, Bolivia o Perú y declaran tener una muy buena situación económica son los que más probabilidades tienen de elegir Europa como destino de un viaje futuro (84 %).

7.2.3.3 Contraste de Hipótesis

Para realizar un contraste de hipótesis definitivo analizaremos la siguiente tabla:

Estimaciones de los parámetros

Viajará Europa ^a	B	Error tít.	Wald	gl	Sig.	Exp(B)	Intervalo de confianza al 95% para Exp(B)	
							Límite inferior	Límite superior
No	Intersección	-1,7	,6	7,8	1	,005		
	[Viajó Chile Bolivia O Perú=,00]	1,4	,5	6,9	1	,009	4,0	1,4 11,2
	[Viajó Chile Bolivia O Perú=1,00]	,0 ^b	.	.	0	.	.	.
	[Economía Muy Buena=,00]	1,0	,5	4,5	1	,035	2,6	1,1 6,5
	[Economía Muy Buena=1,00]	,0 ^b	.	.	0	.	.	.

a. La categoría de referencia es: Si.

b. Este parámetro se ha establecido a cero porque es redundante.

Tabla 7.30

De la tabla anterior, comparando a las personas que viajaron a Chile, Bolivia o Perú contra los que no lo hicieron, se observa el valor mayor a la unidad de e^B ($\text{Exp}(B) = 4,0$), que muestra que es más probable que una persona que ha viajado a esos destinos, realice un viaje a Europa. En términos probabilísticos, una persona que ha viajado a esos destinos tiene 4 veces más chances de realizar un viaje a Europa, que una persona que no lo ha hecho. O sea, las chances de que una persona planea viajar a Europa aumentan 300 % cuando las personas han viajado a Chile, Bolivia o Perú con anterioridad. **Lo que confirma fuertemente la hipótesis 1.**

De la misma manera, comparando a los que declaran tener una muy buena situación económica con los que no lo hicieron, se observa el valor mayor a la unidad de e^B ($\text{Exp}(B) = 2,6$), que muestra que es más probable que las personas de economía muy buena planeen viajar a Europa. En términos probabilísticos, una persona de muy buena situación económica tiene 2,6 veces más chances de planear a viajar a Europa que una persona que no considera tener tan buena situación. O sea, las chances de querer viajar a Europa aumentan un 260 % cuando las personas creen tener una muy buena situación económica. **Lo que confirma fuertemente la hipótesis 2.**

7.2.3.4 Resumen y Conclusiones del modelo

Se obtuvo un modelo que trata de predecir si una persona estaría dispuesta a viajar o no a Europa como destino turístico.

Las cuestiones que resultaron relevantes a la hora de definir esto fueron:

- si la persona en cuestión había viajado a Chile, Bolivia o Perú en el pasado.
- si la persona en cuestión consideraba tener una muy buena situación económica.

Se establecieron dos hipótesis:

Hipótesis 1:

Las personas que han viajado a Chile, Bolivia o Perú en el pasado son más propensas a viajar a Europa en el futuro.

Hipótesis 2:

Las personas que declaran tener una situación económica muy buena son más propensas a viajar a Europa en el futuro.

Se confirmaron ambas hipótesis, ya que se pudo demostrar con significatividad estadística que las personas que viajaron a Chile, Bolivia o Perú presentan un 300 % más chances de elegir viajar a Europa, comparando con los que no lo hicieron. A su vez, las personas que declararon tener una situación económica muy buena

presentan 260 % más chances de elegir Europa como destino, comparando con el resto.

Como caso interesante, se tiene el de las personas que no viajaron a Chile, Bolivia o Perú y que tampoco dicen tener una economía muy buena, con un 34 % de probabilidad de querer viajar a Europa. En oposición, se encuentra el caso de las personas que viajaron a Chile, Bolivia o Perú y también tienen una muy buena situación económica, éstos presentan un 86 % de probabilidades de querer viajar al viejo continente.

7.2.4 Modelo Referente a la Elección de la Modalidad de Organización.

La variable Dependiente en este caso es:

❖ *Modalidad de Organización de Viaje Turístico.*

En este caso se agrupan las modalidades de la siguiente manera, ordenados de menor a mayor independencia:

- Sólo por Agencia de Viajes: En esta categoría se ubican las personas que sólo eligieron la vía más tradicional de organización de un viaje turístico, contactarse con una agencia de viajes, y que esta se ocupe de conseguir y reservar los vuelos (de ser el caso), traslados y el hospedaje.
- Sólo por Internet: En esta categoría se ubican las personas que averiguan o bien reservan tanto vuelos, como traslados u hospedajes a través de la Web.
- Sólo por Cuenta Propia: En esta categoría se encuentran las personas que averiguan lo necesario para sus viajes a través de referencias de amigos, familiares, diarios u otros. En este caso la mayoría de las reservas se realizan por vía telefónica, o bien directamente no se realizan y se emprenden los viajes sin planeamiento previo.

Para la obtención de las variables independientes se siguieron los pasos explicados en el capítulo 6 y las elegidas resultaron:

- *¿Considera su economía mala o regular?*

Esta pregunta toma el valor *Si* cuando el encuestado responde que considera su situación económica mala o regular. Las restantes personas, las que respondieron tener una buena o muy buena economía, se agrupan en *No*.

- *¿Organizó viajes por cuenta propia en el pasado?*

En esta variable se agrupan las personas que han contestado que organizaron viajes en el pasado con la opción *Otros*, en *Si*. Se recuerda que los modos de

organizar un viaje de la encuesta eran con agencia de viajes, por Internet y Otros. En “otros”, los encuestados contestaron casi con exclusividad haber organizado todo lo referente a sus viajes, por cuenta propia. Se permitió a su vez que se responda más de una opción, por lo que una persona, por ejemplo, pudo haber contestado que organizó sus viajes por Internet y también con “otros”, quedando agrupado bajo la respuesta *Si*. Los que no seleccionaron la opción otros, fueron agrupados en *No*.

7.2.4.1 Planteo de Hipótesis

Con respecto a la situación económica de las personas, resulta obvio esperar que las personas con una regular o mala situación tengan más chances de planear a un viaje sin la utilización de una agencia de viajes. Lo contrario, sucede con las personas pudientes.

Por lo tanto:

Hipótesis 1:

Las personas que declaran tener una situación económica buena o muy buena, son más propensas a organizar sus viajes mediante agencias de viaje.

A su vez, es de esperar que las personas que ya han organizado viajes por cuenta propia en oportunidades pasadas, tengan más chances de repetir esto que los demás, o bien optar por una opción intermedia que puede ser Internet. No parece muy probable que su opción predilecta sea la de organizar futuros viajes solamente por agencias de viaje.

Por lo tanto:

Hipótesis 2:

Las personas que han organizado sus viajes por cuenta propia en el pasado, son más propensas a volver a hacerlo, o bien optar por una opción intermedia como puede ser Internet. Este tipo de personas será poco propenso a utilizar solamente agencias de viaje como modalidad de organización.

7.2.4.2 Resultados del Modelo

Resumen del procesamiento de los casos

		N	Porcentaje marginal
Sólo Organizará Sus Viajes	Por Agencia de Viajes	22	26%
	Por Internet	45	53%
	Por Cuenta Propia	18	21%
Economía Mala O Regular	No	62	73%
	Si	23	27%
Organizó Viajes Por Cuenta Propia	No	62	73%
	Si	23	27%
Válidos		85	100%
Perdidos		0	
Total		85	
Subpoblación		4	

Tabla 7.31

En la tabla 7.31 se observa que el 26 % de los encuestados sólo organizarán sus viajes por Agencia, el 53 % sólo lo hará por Internet y el 21 % restante sólo por cuenta propia, totalizando el 100 % de los casos. Además, se observa que el 27 % de los relevados declaran tener una situación económica malo o regular y el 27 % ya organizó viajes por su propia cuenta. Se contemplaron 85 casos.

Con respecto a la clasificación del modelo en cuánto a la elección de la modalidad de organización se presenta la siguiente tabla:

Clasificación

Observado	Pronosticado			Porcentaje correcto
	Por Agencia de Viajes	Por Internet	Por Cuenta Propia	
Por Agencia de Viajes	0	17	5	0%
Por Internet	0	40	5	89%
Por Cuenta Propia	0	5	13	72%
Porcentaje global	0%	73%	27%	62%

Tabla 7.32

Las celdas en diagonal son los casos en que las predicciones son correctas. Se observa que el 89 % de las personas que optan por Internet son bien diagnosticadas y el 72 % de los “cuenta-propistas” se pronostican adecuadamente. El caso de las personas que eligieron agencias de viaje tiene 0 % de aciertos. Como ya hemos explicado en modelos anteriores, esto se debe a que en ninguno de los casos la agencia de viajes es la categoría de respuesta predicha con la probabilidad más alta arrojada por el modelo. De querer segmentar a estas personas se tendrían que buscar otras variables predictoras.

En el global, el 62 % de los casos son bien diagnosticados. Esto supera a cualquier modelo nulo que se quiera plantear.

Con respecto a las frecuencias observadas y pronosticadas:

Frecuencias observadas y pronosticadas

Organizó Viajes Por Cuenta Propia	Economía Mala	O Regular	Sólo Organizará Sus Viajes	Frecuencia			Porcentaje	
				Observada	Pronosticada	Residuo de Pearson	Observada	Pronosticada
No	No	Por Agencia de Viajes	Por Agencia de Viajes	14	14	,1	30%	29%
				32	32	,1	68%	67%
				1	2	-,6	2%	4%
	Si	Por Agencia de Viajes	Por Agencia de Viajes	3	3	-,2	20%	23%
				8	8	-,2	53%	56%
				4	3	,5	27%	22%
Si	No	Por Agencia de Viajes	Por Agencia de Viajes	4	4	-,2	27%	29%
				4	4	-,2	27%	29%
				7	6	,4	47%	42%
	Si	Por Agencia de Viajes	Por Agencia de Viajes	1	1	,5	13%	8%
				1	1	,5	13%	8%
				6	7	-,7	75%	84%

Los porcentajes se basan en las frecuencias observadas totales de cada subpoblación.

Tabla 7.33

Nótese el ajuste satisfactorio del modelo comparando los porcentajes observados y pronosticados.

Con respecto a las hipótesis, nótese que las personas que dicen tener una buena situación económica y no han organizado viajes por cuenta propia, presentan altas probabilidades pronosticadas de organizar por Internet (68 %), menos probabilidades de utilizar agencias de viaje (30 %), y escasas probabilidades pronosticadas de organizar viajes por cuenta propia (2%).

En oposición, se encuentra el caso de las personas de mala situación económica que han organizado viajes por cuenta propia, con pocas probabilidades de contactar una agencia de viajes (13 %), también modestas probabilidades de utilizar Internet (13%) y altas probabilidades de repetir y utilizar su propia gestión (75%).

7.2.4.3 Contraste de Hipótesis

Para realizar un contraste de hipótesis definitivo se analizará la siguiente tabla:

Estimaciones de los parámetros

SóloOrganizaráSusViaje ^a	B	Error típ.	Wald	gl	Sig.	Exp(B)	Intervalo de confianza al 95% para Exp(B)	
							Límite inferior	Límite superior
Por Agencia de Viajes	Intersección	-2,4	,9	7,4	1	,006		
	[EconomíaMala ORegular=,00]	2,0	,8	5,7	1	,017	7,6	1,4 40,1
	[EconomíaMala ORegular=1,00]	,0 ^b	.	.	0	.	.	.
	[OrganizóViajesPor CuentaPropia=,00]	2,4	,8	8,7	1	,003	11,3	2,3 56,0
	[OrganizóViajesPor CuentaPropia=1,00]	,0 ^b	.	.	0	.	.	.
Por Internet	Intersección	-2,3	,8	7,9	1	,005		
	[EconomíaMala ORegular=,00]	2,0	,8	6,4	1	,012	7,2	1,6 33,1
	[EconomíaMala ORegular=1,00]	,0 ^b	.	.	0	.	.	.
	[OrganizóViajesPor CuentaPropia=,00]	3,3	,8	16,9	1	,000	26,4	5,5 125,8
	[OrganizóViajesPor CuentaPropia=1,00]	,0 ^b	.	.	0	.	.	.

a. La categoría de referencia es: Por Cuenta Propia.

b. Este parámetro se ha establecido a cero porque es redundante.

Tabla 7.34

De la tabla 7.34, comparando a las personas de buena situación económica con el resto, se observa el valor mayor a la unidad de e^B ($\text{Exp}(B) = 7,6$), que muestra que es mucho más probable que una persona de buena economía se incline por contactar una agencia de viajes por sobre gestionar sus viajes personalmente, comparando con los de no tan buena situación. En términos estadísticos, una persona mayor de buena situación económica tiene 7,6 veces más chances de utilizar una agencia de viajes, comparando con una persona de no tan buena situación. O sea, las chances de elegir una agencia de viajes por sobre la organización por cuenta propia aumentan 660 % cuando las personas declaran ser pudientes. **Lo que confirma fuertemente la hipótesis 1.**

Algo similar ocurre con la posibilidad de elegir Internet como modalidad de organización. Las chances de elegir Internet por sobre organización por cuenta propia aumentan 620 % cuando las personas declaran no tener mala situación económica.

De la misma manera, de la tabla anterior, comparando a los que han organizado por cuenta propia con los que no lo han hecho, se observa el valor mayor a la unidad de e^B ($\text{Exp}(B) = 11,3$), que muestra que es más probable que los que no son "cuenta-propistas" elijan utilizar una agencia de viajes por sobre gestionar sus viajes por si mismos. En términos estadísticos, se tienen 11,3 veces más chances de elegir una agencia de viajes por sobre propia organización, cuando no se ha organizado por cuenta propia en el pasado. O sea, las chances de optar por una agencia de viajes aumentan un 1030 % cuando las personas no han optado por organizar viajes pasados por cuenta propia.

Algo similar, pero en mayor proporción, sucede cuando se compara Internet con “cuenta propia”. En este caso, las chances de optar por organización vía Internet aumentan un 2540 % cuando las personas no han optado por organizar viajes pasados por cuenta propia.

En este caso sólo se puede confirmar parte de la hipótesis y se rechaza la parte restante. Es cierto que las personas que organizaron viajes por cuenta propia son más propensas a volver a hacerlo en el futuro, pero no es cierto que se inclinarán más a la organización por Internet por sobre la utilización de agencias de viajes.

7.2.4.4 Resumen y Conclusiones del Modelo

Se obtuvo un modelo que trata de predecir entre tres tipos distintos de modalidad de organización de viajes.

Las cuestiones que resultaron relevantes a la hora de definir esto fueron:

- si la persona en cuestión declaraba tener una buena o muy buena situación económica.
- si la persona en cuestión había organizado viajes por cuenta propia o no lo había hecho.

Se establecieron dos hipótesis:

Hipótesis 1:

Las personas que declaran tener una situación económica buena o muy buena, son más propensas a organizar sus viajes mediante agencias de viaje.

Hipótesis 2:

Las personas que han organizado sus viajes por cuenta propia en el pasado, son más propensas a volver a hacerlo, o bien optar por una opción intermedia como puede ser Internet. Este tipo de personas será poco propenso a utilizar solamente agencias de viaje como modalidad de organización.

Se pudo confirmar la primera hipótesis, ya que se pudo demostrar que, a la hora de organizar un viaje, las chances de elegir una agencia por sobre la organización por cuenta propia aumentan 660 % cuando las personas son pudientes.

La segunda hipótesis fue sólo confirmada parcialmente, ya que se pudo probar que las chances de elegir la modalidad “cuenta-propista” aumentan considerablemente cuando ya se ha organizado por esta vía en el pasado. Sin embargo, no es cierto mayoritariamente, que las personas se inclinarán más a la organización por Internet por sobre las agencias de viaje.

Como caso interesante para mencionar del modelo, se encuentra el de las personas de buena situación económica y que no han organizado viajes por cuenta propia, que presentan moderadas probabilidades pronosticadas de contactar una agencia (30 %), más probabilidades de utilizar Internet (68 %), y escasas probabilidades pronosticadas de organizar viajes por cuenta propia (2%).

En oposición, se encuentra el caso de las personas no pudientes, y que han organizado viajes por cuenta propia, con moderadas probabilidades de contactar una agencia de viajes (13 %), también modestas probabilidades de utilizar Internet (13%) y altas probabilidades de repetir y utilizar su propia gestión (75%).

SECCIÓN III

APLICACIONES PRÁCTICAS Y CONCLUSIONES

CAPÍTULO 8 – APLICACIONES PRÁCTICAS DE LOS MODELOS DE REGRESIÓN OBTENIDOS

En esta sección, se evaluarán y compararán las utilidades de distintas alternativas de promoción y publicidad de viajes. Con este fin, se hará uso de la técnica de Simulación de Montecarlo, utilizando muestreos correlacionados (números aleatorios comunes), y en otros casos, se evaluarán alternativas en un árbol de decisión.

Un modelo de simulación estática, a veces llamado simulación de Montecarlo, es el que representa a un sistema en un determinado momento en el tiempo. Es el que se utilizará en nuestro caso, ya que el tiempo no será un elemento en la solución.

El muestreo correlacionado se da cuando se comparan dos sistemas. Significa que, para cada replicación, los mismos números aleatorios son utilizados para simular ambas alternativas. De este modo, la cantidad de replicaciones son iguales para las dos.

La ventaja que tiene esta técnica, de ser bien utilizada, es que se consigue reducir la varianza de las estimaciones puntuales si se compara con el muestro independiente. Una varianza más pequeña implica que el estimador que se consigue será más preciso, ya que se reducen los intervalos de confianza.

Para lograr nuestro objetivo, se cuenta con la información obtenida de los modelos de regresión, y además de información relevada en la Cámara Argentina de Turismo y dos importantes agencias, una de las cuales cuenta con una división exclusiva para turismo juvenil.

8.1 Comparación de Alternativas de envío de Publicidad de Viajes con Hoteles de Lujo para Jóvenes.

Se considera el envío de publicidad a una lista de jóvenes de la base de datos de una agencia de turismo.

Los ingresos supuestos serán los siguientes:

- Ingresos típicos por venta = 640 \$.¹⁷
- Si la venta involucra hoteles de lujo se aplica un coeficiente de 1,6 que multiplica al ingreso típico por venta.

Los costos supuestos serán los siguientes:

- Costo de envío de correo unitario = 14 \$.

¹⁷ Los ingresos para las agencias, en todos los casos, serán por comisiones.

- Costo extra de realizar 2 diseños de publicidad, por unidad = 2 \$.

Otros supuestos serán:

- Cantidad de personas en la base de datos a los que se le envía publicidad = 3500.
- El porcentaje de interesados que finalmente contratan el servicio = 6 %¹⁸
- El desglose de las 3500 personas, en cuanto a datos pasados, será de la siguiente manera:

Viajó hotel	Viajó Chile, Bolivia o Perú	Porcentaje de Población en la muestra
no	no	38%
no	si	9%
si	no	45%
si	si	9%

Tabla 8.1

- Probabilidad de estar interesados en el viaje que se les ofrece: Valores obtenidos de los modelos de regresión de la sección anterior.

8.1.1 Alternativa 1

En una primera aproximación, se enviará a toda esta lista de personas, la oferta de un viaje que incluye un hotel de lujo. Se evalúan en este caso los ingresos, costos y por ende utilidades asociadas a esta alternativa.

En este caso:

¹⁸ La probabilidad de que una persona contrate el servicio se obtiene multiplicando la probabilidad de que esté interesado (obtenida con los modelos de regresión de la sección anterior) por la probabilidad de que finalmente contrate el servicio.

Viajó hotel	Viajó Chile, Bolivia o Perú	Porcentaje de Población en la muestra	Cantidad de Gente de la muestra	Probabilidad de Estar Interesado en Viaje que se le envió en Publicidad	Porcentaje de Interesados que contratan el viaje	Cantidad de Ventas
no	no	38%	1.313	0,29	6%	22
no	si	9%	301	0,09	6%	1
si	no	45%	1.586	0,71	6%	67
si	si	9%	301	0,36	6%	6

Tabla 8.2

En la tabla 8.2, el caso de las personas que no viajaron en el pasado a hoteles de lujo ni a Chile, Bolivia o Perú consiste en el 38 % de los integrantes de la base, 1313 personas. De esta cantidad, el 29 % estarán interesados en la publicidad que les fue enviada¹⁹. Finalmente, el 6 % de esos interesados contratarán el viaje, lo que constituye 22 ventas. Para el resto de las personas de la base, el razonamiento es análogo.

Ahora bien, veamos los ingresos, costos y utilidades:

Viajó hotel	Viajó Chile, Bolivia o Perú	Cantidad de Ventas	Ingresos Por Comisiones de Ventas	Costo de Envío de Publicidad	Utilidad
no	no	22	22.528	18.375	4.153
no	si	1	1.024	4.211	-3.187
si	no	67	68.608	22.203	46.405
si	si	6	6.144	4.211	1.933

Ut. Total Alternativa 1	49.304
-------------------------	--------

Tabla 8.3

Nótese, que si bien la utilidad total²⁰ de esta alternativa es positiva, el caso de las personas que no viajaron a hoteles de lujo en el pasado y que sí viajaron a Chile, Bolivia o Perú es un rubro deficitario, ya que los costos asociados de envío de publicidad superan a los ingresos por comisiones.

8.1.2 Alternativa 2

Ya que existe un grupo de personas, al cual no resulta rentable enviarle publicidad de viajes que incluyan hoteles de lujo, se considera la posibilidad de no enviarles ningún tipo de publicidad a esas personas. En ese caso los resultados serían los siguientes:

¹⁹ Dato obtenido de los modelos de regresión de la sección anterior.

²⁰ Vale aclarar que en todos los casos que se explicarán, no se consideran gastos de administración, financiación ni fiscales

Viajó hotel	Viajó Chile, Bolivia o Perú	Cantidad de Ventas	Ingresos Por Comisiones de Ventas	Costo de Envío de Publicidad	Utilidad
no	no	22	22.528	18.375	4.153
no	si				
si	no	67	68.608	22.203	46.405
si	si	6	6.144	4.211	1.933
Ut. Total Alternativa 2					52.491

Tabla 8.4

Por lo tanto, la variación de utilidad es la siguiente:

Variación de Utilidad Alt 2 vs Alt 1	3.187
--------------------------------------	--------------

Tabla 8.5

Ahora bien, aunque la segunda alternativa aparente arrojar mejores resultados, existen costos no visibles asociados al no envío de publicidad a ciertas personas. Estos, de enterarse que no han sido tenidas en cuenta, podrían no sentirse a gusto con la agencia, y buscarse otra en la que se sientan más importantes. Una alternativa que surge ante esta situación, es la de envío de una publicidad diferencial a estas personas, que involucre un viaje en el cual no se ofrezcan hoteles de lujo. Es más, la agencia podría verse favorecida, y establecer una administración de Marketing en la que se promociona como una agencia que ofrece paquetes a medida, según el perfil del viajante.

8.1.3 Alternativa 3

De esta manera, surge esta nueva alternativa, en la que se incurre en un costo extra ya que se realizarán diseños extra de publicidad, uno es el que se utilizó en las alternativas 1 y 2, y el otro en el que se ofrecen hospedajes de menor nivel. A raíz de este nuevo costo se espera que los ingresos sean mayores.

En esta instancia, se enviará a las personas con probabilidades menores al 50 % de contratar un hotel de lujo (a tres de los cuatro grupos que se marcan en amarillo en la tabla 8.6), la oferta de un viaje que incluye alojamiento de menor nivel. Por lo tanto, las probabilidades de contratación de servicios aumentan, así como la cantidad de ventas:

Viajó hotel	Viajó Chile, Bolivia o Perú	Porcentaje de Población en la muestra	Cantidad de Gente de la muestra	Probabilidad de Estar Interesado en Viaje que se le envió en Publicidad	Porcentaje de Interesados que contratan el viaje	Cantidad de Ventas
no	no	38%	1.313	0,71	6%	55
no	si	9%	301	0,91	6%	16
si	no	45%	1.586	0,71	6%	67
si	si	9%	301	0,64	6%	11

Tabla 8.6

Se recuerda que si bien hay más ventas, éstas son de menor categoría y por ende menos rentables. A su vez, los costos de envío de publicidad son mayores:

Viajó hotel	Viajó Chile, Bolivia o Perú	Cantidad de Ventas	Ingresos Por Comisiones de Ventas	Costo de Envío de Publicidad	Utilidad
no	no	55	35.200	21.000	14.200
no	si	16	10.240	4.813	5.428
si	no	67	68.608	25.375	43.233
si	si	11	7.040	4.813	2.228

Ut. Total Alternativa 3	65.088
-------------------------	--------

Tabla 8.7

8.1.4 Comparación de Resultados

De esta manera, la comparación de utilidades resulta:

Variación de Utilidad Alt 3 vs Alt 2	12.597
--------------------------------------	--------

Variación de Utilidad Alt 3 vs Alt 1	15.784
--------------------------------------	--------

Tabla 8.8

Ahora si, la tercer alternativa parece la más viable tanto en términos de Marketing como económicamente, al incluir a todos los integrantes de la base y tener las mejores utilidades.

El paso siguiente es introducir variabilidad en los costos, ingresos y supuestos que se han considerado, con el fin de simular varios ensayos, para que el análisis tenga mayor validez. Se hará uso del Software Crystal Ball y de la técnica de los números aleatorios comunes.

Las distribuciones asociadas a los distintos supuestos serán las siguientes:

En cuanto a los ingresos:

- Ingresos típicos por venta: Distribución normal con media de \$640 y desvío estándar de 10% de la media.
- Coeficiente que multiplica al ingreso típico por venta de concurrir a un hotel de lujo: Distribución triangular²¹ con valor mínimo de 1,44, valor más probable de 1,6 y valor máximo de 1,76.

En cuanto a los costos:

²¹ La distribución triangular no existe en la naturaleza, se utiliza básicamente en evaluación de proyectos.

- Costo de envío de correo unitario: Distribución triangular con valor mínimo de \$12, valor más probable de \$14 y valor máximo de \$16.
- Costo extra de realizar diseños extras, por unidad: Distribución triangular con valor mínimo de \$1,5, valor más probable de \$2 y valor máximo de \$2,5.

Otros supuestos serán:

- Porcentaje de interesados que finalmente contratan el servicio: Distribución triangular con valor mínimo de 4%, valor más probable de 6% y valor máximo de 8%.
- Probabilidad de estar interesados en el viaje que se les ofrece: Distribución normal con media extraída de los modelos de regresión obtenidos en cada caso y desvío estándar de 20% de la media.

Los resultados fueron los siguientes:

Comparación alternativa 2 versus alternativa 1

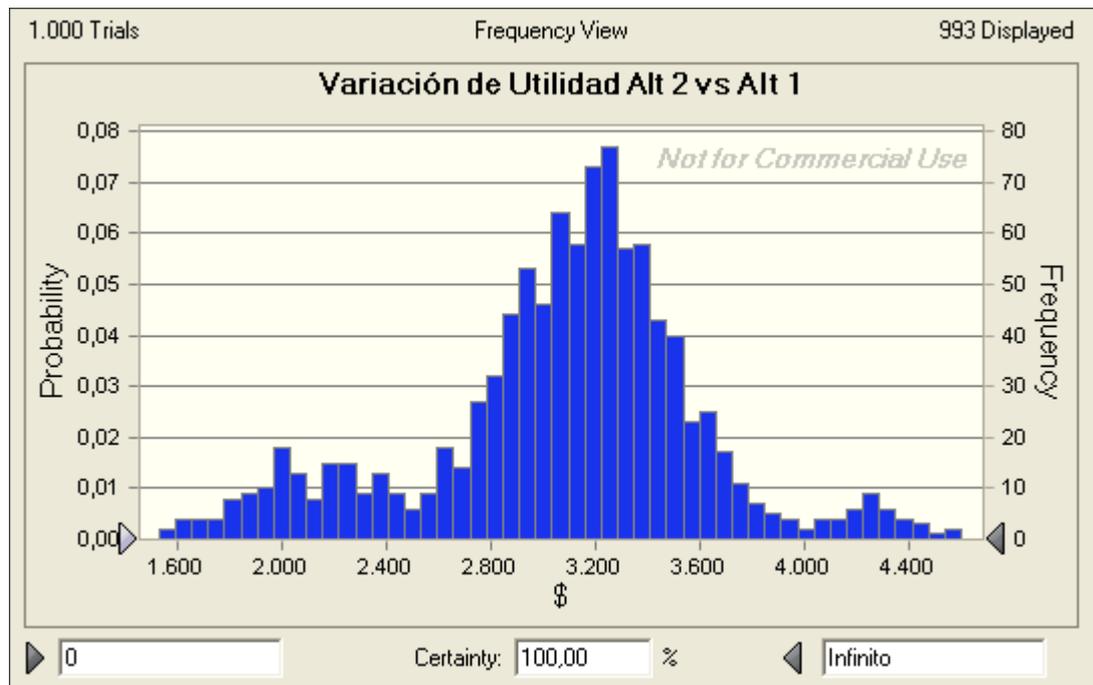


Gráfico 8.1

Como se observa en el gráfico, en el 100 % de los ensayos la alternativa 2 tuvo mejores utilidades que la alternativa 1.

Estimador puntual : $\bar{D} = 3.068$

Varianza de la muestra : $S_D^2 = 300.383$

Desv. Estd. : $s.e.(\bar{D}) = S_D / \sqrt{R} = 548 / \sqrt{1000} = 17$

Grados de Libertad : $\nu = R - 1 = 999$

95% c. i. para $\theta_2 - \theta_1$: con $t_{\alpha/2, \nu} = t_{0.025, 999}$

$$3.034 \leq \theta_2 - \theta_1 \leq 3.102$$

De esta manera, ya que el intervalo de confianza de la resta de utilidades de las alternativas no contiene el cero, se demuestra que la alternativa 2 es superior en término de utilidades a la alternativa 1.

Ahora bien, incorporando a la alternativa 3 a la comparación obtenemos:

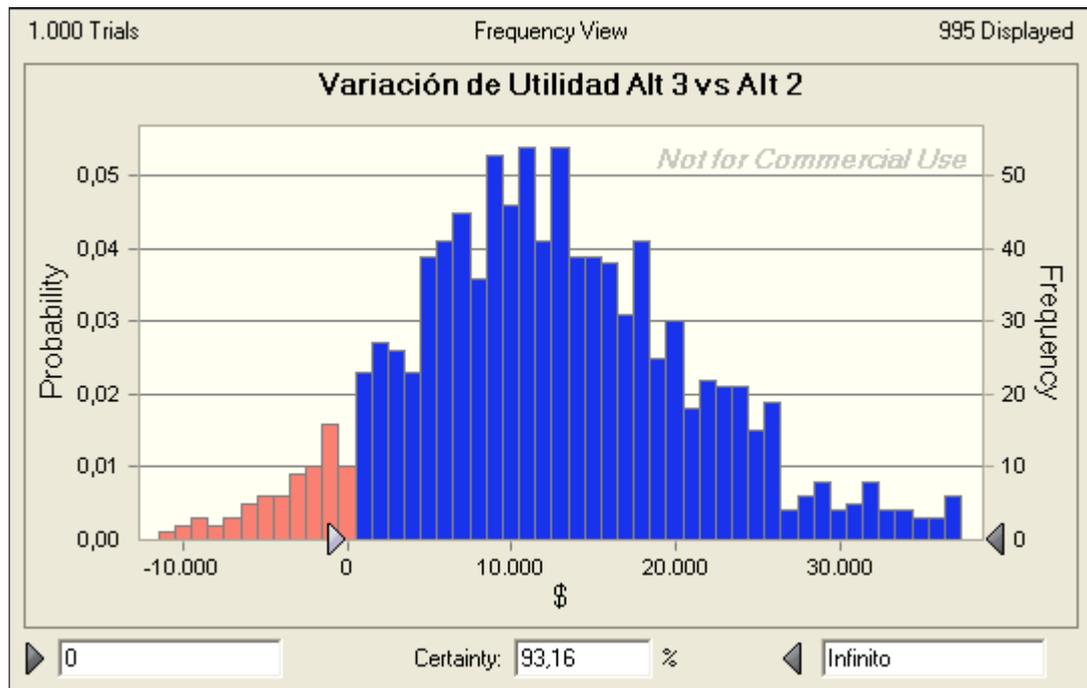


Gráfico 8.2

Como se observa en el gráfico 8.2, en el 93 % de los ensayos la alternativa 3 tuvo mejores utilidades que la alternativa 2.

$$\text{Estimador puntual : } \bar{D} = 12.473$$

$$\text{Varianza de la muestra : } S_D^2 = 79.343.085$$

$$\text{Desv. Estd. : } s.e.(\bar{D}) = S_D / \sqrt{R} = 8907 / \sqrt{1000} = 282$$

$$\text{Grados de Libertad : } \nu = R - 1 = 999$$

$$95\% \text{ c. i. para } \theta_3 - \theta_2 : \text{ con } t_{\alpha/2, \nu} = t_{0.025, 999}$$

$$11.921 \leq \theta_2 - \theta_1 \leq 13.025$$

siendo θ_i la utilidad promedio de la alternativa i .

Concluyendo, ya que el intervalo de confianza de la resta de utilidades de las alternativas no contiene el cero, se demuestra que la alternativa 3 es superior en término de utilidades a la alternativa 2, y es la mejor alternativa en el global. Supera en un 31% a la alternativa 1 y en un 23 % a la alternativa 2, en valores esperados de utilidad. A su vez, es un 16 % más rentable²² que la alternativa 1.

De esta forma, la mejor estrategia a seguir parece ser la de enviar a las personas con probabilidades menores al 50 % de contratar un hotel de lujo (a tres de los cuatro grupos), la oferta de un viaje que incluya alojamiento de menor nivel. Al resto de las personas, sí se le envía ofertas que involucren hoteles de categoría.

8.2 Comparación de Alternativas de envío de Publicidad de Viajes con Distintos Destinos para Jóvenes.

Se considera el envío de publicidad a una lista de jóvenes de la base de datos de una agencia de turismo. El proyecto excluye la oferta de destinos locales.

Los ingresos supuestos serán los siguientes:

- Ingresos típicos por venta (Región 2: Países Limítrofes²³) = 1110 \$.
- Si la venta involucra viajes al *Resto del mundo*²⁴ (Región 3) se aplica un coeficiente de 1,2 que multiplica al ingreso típico por venta.

²² Rentabilidad en este caso será el cociente entre utilidad y costos de envío de publicidad.

²³ Se recuerda que “países limítrofes” se considera: Uruguay, Brasil, Bolivia, Chile, Perú.

²⁴ Se recuerda que “resto del mundo” se considera: Colombia, Centroamérica, Norteamérica, Oceanía, África o Asia

Los costos supuestos serán los siguientes:

- Costo de envío de correo unitario = 14 \$.
- Costo extra de realizar 2 diseños de publicidad, por unidad = 2 \$.

Otros supuestos serán:

- Cantidad de personas en la base de datos a los que se le envía publicidad = 3500.
- El porcentaje de interesados que finalmente contratan el servicio = 8 %
- El desglose de las 3500 personas, en cuanto a datos pasados, será de la siguiente manera:

Viajó Resto de América	Porcentaje de Población en la muestra
no	78%
si	22%

- Probabilidad de estar interesados en el viaje que se les ofrece: Valores obtenidos de los modelos de regresión de la sección anterior.

8.2.1 Alternativa 1

En una primera aproximación, se enviará a toda esta lista de personas, la oferta de un viaje que tenga el destino *Resto del Mundo*. Se evalúan en este caso los ingresos, costos y por ende utilidades asociadas a esta alternativa.

En este caso:

Viajó Resto de América	Porcentaje de Población en la muestra	Cantidad de Gente de la muestra	Probabilidad de Estar Interesado en Viaje que se le envió en Publicidad	Porcentaje de Interesados que contratan el viaje	Cantidad de Ventas
no	78%	2734	0,33	8%	72
si	22%	766	0,679	8%	41

Tabla 8.9

En la tabla 8.9, el caso de las personas que no viajaron en el pasado al resto de América consiste en el 78 % de los integrantes de la base, 2734 personas. De

esta cantidad, el 33 % estarán interesados en la publicidad que les fue enviada²⁵. Finalmente, el 8 % de esos interesados contratarán el viaje, lo que constituye 72 ventas. Para el resto de las personas de la base, el razonamiento es análogo.

Ahora bien, veamos los ingresos, costos y utilidades:

Viajó Resto de América	Porcentaje de Población en la muestra	Cantidad de Ventas	Ingresos Por Comisiones de Ventas	Costo de Envío de Publicidad	Utilidad
no	78%	72	95.904	38.281	57.623
si	22%	41	54.612	10.719	43.893
Ut. Total Alternativa 1					101.516

Tabla 8.10

Nótese que, a diferencia de la alternativa 1 del caso de los hoteles de lujo, no hay ningún rubro deficitario, ya que los costos asociados de envío de publicidad no superan a los ingresos por comisiones en ninguno de los dos grupos de personas.

8.2.2 Alternativa 2

Ya que resulta rentable enviarle publicidad de viajes que incluyan como destino el *resto del mundo* a ambos grupos, la posibilidad de no enviar ningún tipo de publicidad a un grupo determinado, sólo tendría sentido de tener un presupuesto limitado. En ese caso, por ejemplo, se les podría enviar publicidad a las personas que sí han viajado al Resto de América, ya que sería el más lucrativo. Los resultados serían los siguientes:

Viajó Resto de América	Porcentaje de Población en la muestra	Cantidad de Ventas	Ingresos Por Comisiones de Ventas	Costo de Envío de Publicidad	Utilidad
no	78%				
si	22%	41	54.612	10.719	43.893
Ut. Total Alternativa 2					43.893

Tabla 8.11

La diferencia de utilidades entre alternativas es abrumadora, por lo que no haría falta un análisis de comparación de medias formal como se hizo en casos anteriores.

8.2.3 Alternativa 3

²⁵ Dato obtenido de los modelos de regresión de la sección anterior.

Ahora bien, en lugar de no enviar ningún tipo de publicidad al grupo que no ha viajado al resto de América, surge la posibilidad de enviarles ofertas de viajes en la que el destino sea un país limítrofe. De este modo, se incurre en un costo extra ya que se realizarán más diseños, unos son los que fueron utilizados en las alternativas 1 y 2, y los nuevos en los que se ofrece esta nueva posibilidad. A raíz de este nuevo costo, se espera que los ingresos sean mayores.

A raíz de estos, la probabilidad de estar interesado en la oferta de uno de los dos grupos aumenta, pasando de 33 % a 45 %. Por ende, aumentarán las ventas como se ve en la tabla 8.12:

Viajó Resto de América	Porcentaje de Población en la muestra	Cantidad de Gente de la muestra	Probabilidad de Estar Interesado en Viaje que se le envió en Publicidad	Porcentaje de Interesados que contratan el viaje	Cantidad de Ventas
no	78%	2734	0,45	8%	98
si	22%	766	0,679	8%	41

Tabla 8.12

Se recuerda que si bien hay más ventas, éstas son de menor categoría y por ende menos rentables. A su vez, los costos de envío de publicidad son mayores:

Viajó Resto de América	Porcentaje de Población en la muestra	Cantidad de Ventas	Ingresos Por Comisiones de Ventas	Costo de Envío de Publicidad	Utilidad
no	78%	98	108.780	43.750	65.030
si	22%	41	54.612	12.250	42.362

	107.392
--	----------------

Tabla 8.13

8.2.4 Comparación de resultados

De esta manera, la comparación de utilidades resulta:

Variación de Utilidad Alt 3 vs Alt 1	5.876
--------------------------------------	--------------

Tabla 8.14

Ahora si, la tercer alternativa parece la mas viable tanto en términos de Marketing como económicamente, al incluir a todos los integrantes de la base y tener las mejores utilidades. Se comparan solamente las alternativas 1 y 3, ya que como se dijo con anterioridad, la alternativa 2 sólo tiene sentido en algún caso particular.

El paso siguiente es introducir variabilidad en los costos, ingresos y supuestos que se han considerado, con el fin de simular varios ensayos, para que el análisis tenga mayor validez. Se hará uso del Software Crystal Ball y de la técnica de los números aleatorios comunes.

Las distribuciones asociadas a los distintos supuestos serán las siguientes:

En cuanto a los ingresos:

- Ingresos típicos por venta: Distribución normal con media de \$1110 y desvío estándar de 10% de la media.
- Coeficiente que multiplica al ingreso típico por venta de concurrir al *resto del mundo*: Distribución triangular con valor mínimo de 1,1, valor más probable de 1,2 y valor máximo de 1,3.
- Coeficiente que multiplica al ingreso típico por venta de concurrir países limítrofes: Distribución triangular con valor mínimo de 0,9, valor más probable de 1 y valor máximo de 1,1.

En cuanto a los costos:

- Costo de envío de correo unitario: Distribución triangular con valor mínimo de \$12, valor más probable de \$14 y valor máximo de \$16.
- Costo extra de realizar diseños extras, por unidad: Distribución triangular con valor mínimo de \$1,5, valor más probable de \$2 y valor máximo de \$2,5.

Otros supuestos serán:

- Porcentaje de interesados que finalmente contratan el servicio: Distribución triangular con valor mínimo de 6%, valor más probable de 8% y valor máximo de 10%.
- Probabilidad de estar interesados en el viaje que se les ofrece: Distribución normal con media extraída de los modelos de regresión obtenidos en cada caso y desvío estándar de 20% de la media.

Los resultados fueron los siguientes:

Comparación alternativa 2 versus alternativa 1

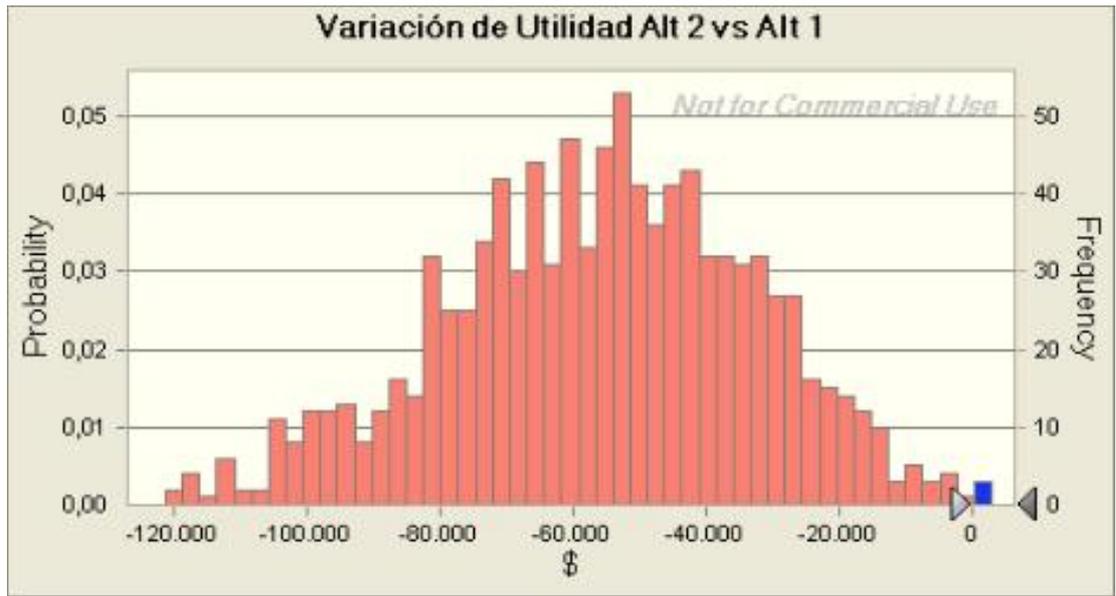
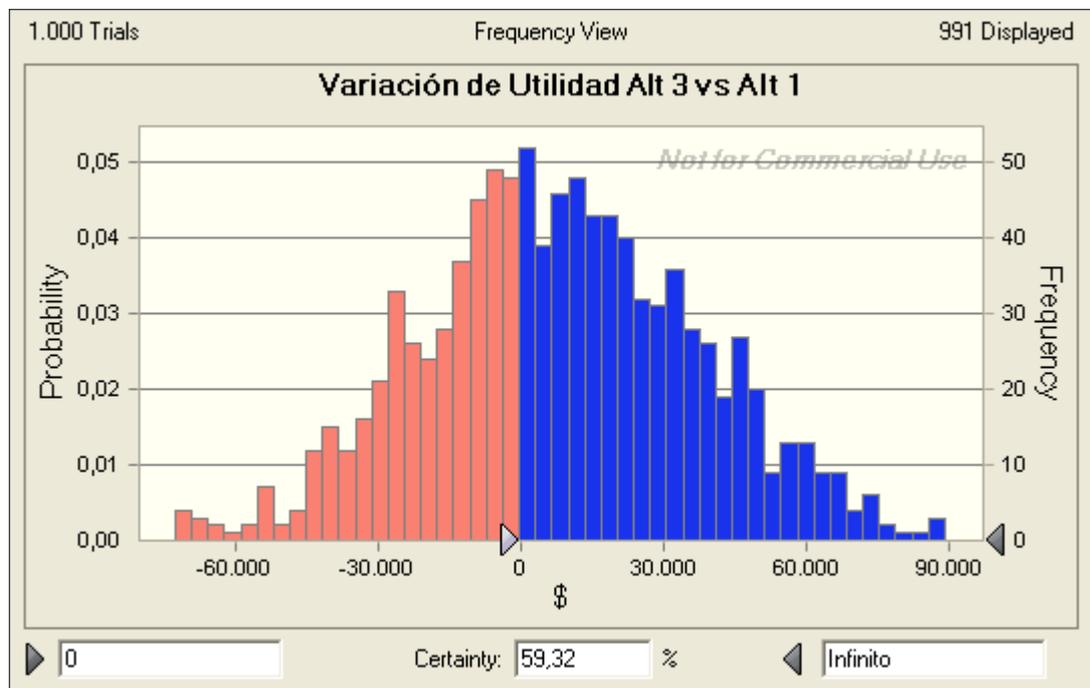


Gráfico 8.3

En el gráfico anterior, sólo en el 1,6 % de los ensayos la alternativa 2 tuvo mejores utilidades que la alternativa 1.

Ahora bien, la comparación que nos interesa será la de la alternativa 3 contra la alternativa 1:



Como se observa en el gráfico, en el 59,32 % de los ensayos la alternativa 3 tuvo mejores utilidades que la alternativa 1.

Comparando las utilidades en 1000 ensayos:

$$\text{Estimador puntual : } \bar{D} = 7.592$$

$$\text{Varianza de la muestra : } S_D^2 = 907.082.690$$

$$\text{Desv. Estd. : } s.e.(\bar{D}) = S_D / \sqrt{R} = 30.118 / \sqrt{1000} = 952$$

$$\text{Grados de Libertad : } \nu = R - 1 = 999$$

95% c. i. para $\theta_3 - \theta_2$: con $t_{\alpha/2, \nu} = t_{0.025, 999}$

$$5.725 \leq \theta_2 - \theta_1 \leq 9.459$$

siendo θ_i la utilidad promedio de la alternativa i .

Concluyendo, ya que el intervalo de confianza de la resta de utilidades de las alternativas no contiene el cero, se demuestra que la alternativa 3 es superior en término de utilidades a la alternativa 1.

De esta forma, la mejor estrategia estará ligada a los costos en los que se puede y quiere incurrir. La alternativa de menor costo sería la 2, que a su vez es la más rentable. De poseer más presupuesto, se podría optar entre las opciones 3 y la 1. La alternativa 3 mejora las utilidades esperadas en un 6 % con respecto a la alternativa 1, pero los costos asociados de esta posibilidad, aumentan un 14 %.

8.3 Comparación de Alternativas de Publicidad para Distintas Modalidades de Organización de Viajes

Recordando de los modelos obtenidos en la sección anterior, las modalidades de organización de viajes estudiadas en este trabajo eran:

- por contacto con una Agencia de Viajes

Supongamos ser parte del área comercial de una agencia de viajes, entonces nos interesará contactar a la gente que esté más ávida a utilizar esta

modalidad, ya que probablemente contraten paquetes con servicios muy rentables.

- por Internet

En este caso, aunque a simple vista no se note, las agencias también tienen incidencia. Muchas agencias tienen página Web, y ofrecen servicios por esta vía. Lo que sí se aclara, es que en general las personas que utilizan la Web son más “independientes” y es probable que contraten uno o un par de servicios por esta vía, pero no paquetes enteros.

- por Cuenta Propia

Ahora sí, las agencias no tendrían ninguna incidencia en este caso. Dentro de una base de datos de personas, sería buena idea tratar de evitar invertir en estos. De aquí la importancia de los modelos obtenidos en la sección anterior, ya que se pudo realizar una correcta segmentación, individualizando las características que poseen estos individuos.

La idea en esta oportunidad será diseñar una campaña de publicidad que implique el mayor valor económico posible, en este caso, para una agencia de viajes. Para tal fin, se evaluarán distintas opciones que luego serán evaluadas en un árbol de decisión.

La campaña será destinada a público joven, de menos de 30 años de edad. Supongamos contar con una base de datos de 8000 personas²⁶, que se segmentan de esta manera:

Organizó Por cuenta Propia	Economía Mala o Regular	Porcentaje de Población	Probabilidad de uso de Agencia	Probabilidad de uso de Internet
no	no	55%	29%	67%
no	si	18%	23%	56%
si	no	18%	27%	29%
si	si	9%	8%	8%

Tabla 8.14

Con esta base se evaluarán las opciones que se explicarán a continuación:

8.3.1 Opción A

Esta opción considera el envío de publicidad de viajes ofrecidos por la agencia a todas las personas de la base. Los ingresos, costos y suposiciones asociadas son:

Ingresos:

²⁶ Se aclara que esta no es una base de datos de clientes de una agencia sino de personas en general.

Ingreso por venta \$ 1.000

Costos:

Costo envío Publicidad para Agencia/Unidad \$ 11

Suposiciones:

Probabilidad de contratar el servicio de estar interesado - Optimista 6%

Probabilidad de contratar el servicio de estar interesado - Pesimista 3%

En cuanto a las suposiciones, al no conocer ciertamente las probabilidad de éxito de esta opción, se establecen dos escenarios equiprobables: uno de ellos optimista en el que el 6 % de los interesados finalizan contratando servicios, y otro pesimista en la que el 3 % de los interesados terminan haciéndolo.

De esta manera, con la base que poseemos los resultados en ventas serían los siguientes:

Organizó Por cuenta Propia	Economía Mala o Regular	Cantidad de Ventas Optimista	Cantidad de Ventas Pesimista
no	no	77	38
no	si	19	10
si	no	23	11
si	si	4	2

Tabla 8.15

Ahora bien, las utilidades que surgen del ingreso por venta por la cantidad de ventas, restando los costos de envío serán los siguientes:

Utilidad Optimista	Utilidad Pesimista
\$ 34.936	-\$ 26.532

Como los dos escenarios son equiprobables, la utilidad de la alternativa surge del promedio entre estos dos resultados exhibidos:

Utilidad de la Alternativa
\$ 4.202

Tabla 8.16

8.3.2 Opción B

En este caso se podría dejar de mandar publicidad al grupo de personas que resultan poco rentables para nuestro propósito. Es el caso de las personas que no tienen una buena economía y son "cuenta propistas". De esta manera se ahorran 7.920 \$ de envíos de publicidad y se resignan cuatro ventas en el caso optimista, y 2 en el caso pesimista.

Las utilidades asociadas serían:

Utilidad	Utilidad
Optimista	Pesimista
\$ 39.605	-\$ 20.056

Al ser equiprobables:

Utilidad de la Alternativa
\$ 9.774

Tabla 8.17

Nótese como han mejorado notablemente las utilidades, con este simple cambio.

8.3.3 Opción C

Esta opción consiste en publicitar en diferentes páginas de Internet. Los ingresos, costos y suposiciones asociadas son:

Ingresos:

Ingreso por venta en Internet	\$ 450
-------------------------------	--------

Los ingresos en este caso, son menores por cada venta, ya que en Internet es muy común la compra de, por ejemplo, transportes aéreos o city tours, que no son tan rentables para las agencias pues no es lo mismo que vender paquetes completos.

Costos:

Costo Publicación Internet \$ 25.000

Estos costos se asocian a publicaciones en páginas Web de importante repercusión.

Suposiciones:

Probabilidad de contratar el servicio de estar interesado - Optimista 0,80%

Probabilidad de contratar el servicio de estar interesado - Pesimista 0,20%

Lógicamente, las probabilidades de una concretar una venta son mucho menores en este caso, comparando con el envío de publicidad a cada persona de las opciones a y b.

En este caso, no hay forma directa de discriminar a quién llega la publicidad y a quién no, como en los casos anteriores.

Las ventas serían las siguientes:

Organizó Por cuenta Propia	Economía Mala o Regular	Cantidad de Ventas Optimista	Cantidad de Ventas Pesimista
no	no	88	22
no	si	28	7
si	no	28	7
si	si	15	4

Tabla 8.18

Las utilidades resultarían:

Utilidad Optimista	Utilidad Pesimista
\$ 47.000	-\$ 7.000

Al ser estas dos posibilidades equiprobables la utilidad resultante es:

Utilidad de la Alternativa
\$ 20.000

Tabla 8.19

8.3.4 Opción D

En este caso se plantea la posibilidad de realizar las opciones B y C juntas. Se elige la B por sobre la A ya que se ha demostrado ser más rentable. Así pues, no se le envía publicidad de viajes a los que sólo tienen 8 % de elegir esta modalidad, a todo el resto sí se les envía, y además se realizan las publicaciones en Internet.

De este modo, se abren cuatro escenarios, que surgen de la combinación de los escenarios planteados anteriormente. De nuevo, éstos serán equiprobables:

Escenario	Probabilidad Asignada
Opt A - Opt I	25%
Opt a - Pes I	25%
Pes A - Opt I	25%
Pes A - Pes I	25%

Tabla 8.20

Opt A – Opt I: se refiere al escenario en el cual las alternativas B y C, son ambas exitosas.

Opt A – Pes I: se refiere al escenario en el cual la alternativa B es exitosa y la alternativa C no resulta de esta manera.

Pes A – Opt I: se refiere al escenario en el cual la alternativa C es exitosa y la alternativa B no resulta de esta manera.

Pes A – Pes I: se refiere al escenario en el cual las alternativas B y C, no resultan exitosas.

Las utilidades asociadas serían:

UT Opt A - Opt I	UT Opt a - Pes I	UT Pes A - Opt I	UT Pes A - Pes I
\$ 86.605	\$ 32.605	\$ 26.944	-\$ 27.056

Tabla 8.21

La utilidad de la alternativa en el global sería:

Utilidad de la Alternativa
\$ 29.774

Tabla 8.22

La utilidad de esta alternativa coincide con la suma de las alternativas de las utilidades de las alternativas B y C.

8.3.5 Opción E

Se podría reflexionar acerca de la opción D: ¿No sería mejor realizar pruebas pilotos antes de arriesgarse a que los escenarios sean pesimistas? Hasta ahora hemos hecho caso omiso de un aspecto extremadamente importante de muchas de las oportunidades (si no es la mayoría) de inversión: la capacidad de los administradores de demorar el inicio de un proyecto o, una vez iniciado, de expandirlo o cancelarlo. Si no se toman en cuenta estas opciones *reales* se puede ocasionar que se subestime el valor del proyecto.²⁷

Por lo pronto, esta opción plantea el envío de publicidad como en la opción B, pero sólo a un 10 % de las personas de la base. A su vez, en paralelo, sugiere que también se realice una publicación en Internet (pero de menor envergadura que en la opción C), que sólo involucre 2.000 personas aproximadamente. Los costos asociados serían proporcionales en el caso del envío de publicidad, y el costo de la publicación piloto en Internet sería de \$ 3.000.

De esta manera, de acuerdo al éxito o fracaso de estas dos pruebas pilotos, se procederá a elegir entre las cuatro opciones que surgen:

- Opción 1: No envío publicidad de agencia a los que solo tienen 8 % de probabilidad de elegir esta modalidad, a todo el resto sí les envío y además publico en Internet (sería la opción D).
- Opción 2: No envío publicidad de agencia a los que solo tienen 8 % de probabilidad de elegir esta modalidad, a todo el resto sí les envío (sería la opción B).
- Opción 3: Realizar la publicación completa en Internet (sería la opción C).
- Opción 4: Suspender las acciones por completo (esta nueva opción surge ya que estamos considerando opciones *reales*).

Supongamos, para una mayor comprensión de esta opción, que la prueba piloto del envío de publicidad no resulta exitosa y sí la publicación piloto en Internet.

De esta manera, los resultados de las opciones a elegir serían las siguientes²⁸:

- Opción 1: Utilidades = \$25.596
- Opción 2: Utilidades = \$-20.057
- Opción 3: Utilidades = \$44.650

²⁷ Adaptado de: "Finanzas" de Bodie y Merton, a partir de la página 448.

²⁸ Se aclara que en el caso en que las opciones 1, 2, 3 o 4 resulten positivas, se aplica un coeficiente de 0,95 a éstas como penalidad, ya que se demoran en el tiempo las ejecuciones de estas opciones mientras se hacen las pruebas pilotos.

- Opción 4: Utilidades = \$0

Se observa que la mejor opción es la 3 y, por lo tanto, de no resultar exitosa la prueba piloto del envío de publicidad y sí la publicación piloto en Internet, la opción a elegir sería suspender el envío de publicidades y realizar la publicación completa en Internet.

Para una visión global de todo el proyecto, se muestra el siguiente árbol de decisión, en el gráfico 8.5, que define las mejores opciones y resultados en cada caso:

En verde se coloca la mejor opción, que es la elegida en cada caso
En rojo se colocan las opciones descartadas

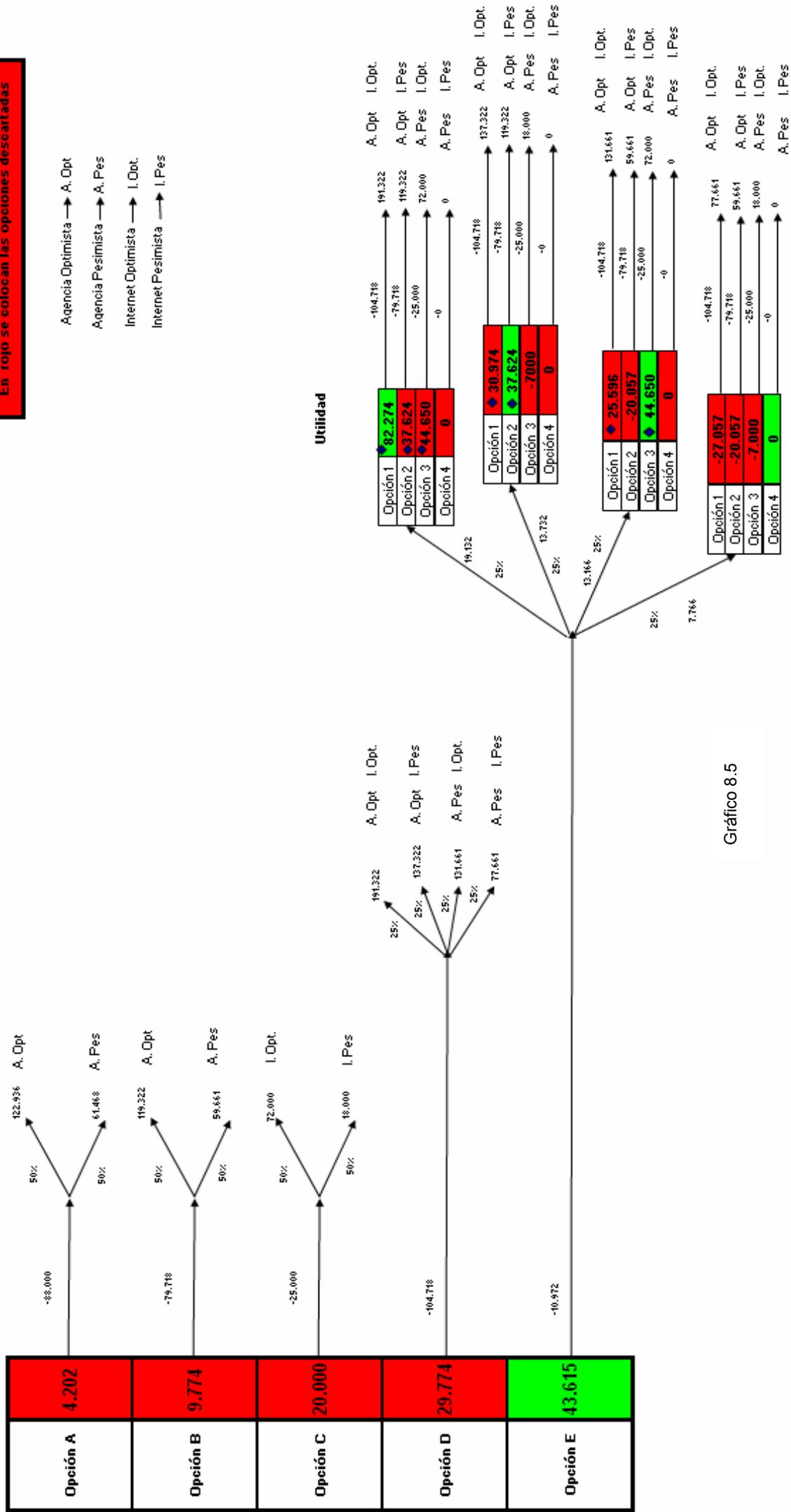


Gráfico 8.5

Como se pudo observar en el árbol de decisión, la mejor opción resulta la E, ya que supera en un 46 % en término de utilidades a la opción D, que es la segunda mejor opción.

Por lo tanto, los pasos a seguir serían realizar ambas pruebas pilotos y elegir la mejor opción en cada caso. De esta forma, se aseguran los mejores valores económicos del proyecto.

CAPÍTULO 9 – RESUMEN Y CONSIDERACIONES FINALES

Acorde a los objetivos de esta investigación, se han logrado diseñar modelos predictivos que sirven de apoyo a las decisiones comerciales adoptadas en distintos oferentes del rubro turismo, basados en un trabajo de campo que fue diseñado y realizado en su totalidad para este proyecto. Se aplicó, entre otras, la técnica estadística de regresión logística multinomial, por ser ésta la que mejor se adaptaba al análisis del tipo de datos manejados en el estudio.

Se dividieron los modelos en dos partes: unos referentes a la estimación de probabilidades de ocurrencia de sucesos relacionados con el turismo emisor, con respecto a población en general, y otros análogos pero basándose sólo en población joven. Todos los modelos obtenidos resultaron altamente significativos, lo que refuerza la validez del análisis realizado.

Así pues se obtuvieron conclusiones importantes acerca de tipo de hospedaje predilecto, de destinos preferidos por ciertas personas, y de modalidades de organización de viajes.

9.1 Sumario de los modelos generales obtenidos

- Con respecto al tipo de hospedaje:

Las personas casadas o en pareja presentan un 275 % más chances de elegir hoteles de lujo en un viaje futuro, comparando con los solteros. A su vez, las personas que ya han utilizado este tipo de alojamiento en el pasado presentan un 490 % más chances de volver a utilizarlo en el futuro, comparando con los que nunca lo hicieron.

- Con respecto a los destinos:

Las personas que viajaron al resto de América en el pasado, presentan un 163 % más chances de viajar al *resto del mundo* en un futuro viaje, comparando contra los que no viajaron. A su vez, las personas que no declaran un gasto modesto en turismo presentan un 145 % más chances de viajar al *resto del mundo*, comparando con los que declaran un gasto modesto.

Las personas que viajaron a Chile, Bolivia o Perú presentan un 275 % más chances de elegir viajar a Europa, comparando con los que no lo hicieron. A su vez, las personas que declararon tener una situación económica muy buena presentan 200 % más chances de elegir Europa como destino, comparando con el resto.

- Con respecto a la modalidad de organización:

Las chances de elegir una agencia de viajes por sobre la organización por cuenta propia aumentan 590 % cuando las personas tienen más de 30 años.

9.2 Sumario de los modelos obtenidos para gente joven

- Con respecto al tipo de hospedaje:

Las personas que no viajaron a Chile, Bolivia o Perú presentan un 319 % más chances de elegir hoteles de lujo en un viaje futuro, comparando con los que sí lo hicieron. A su vez, las personas que ya han utilizado este tipo de alojamiento en el pasado presentan un 490 % más chances de volver a utilizarlo en el futuro, comparando con los que nunca lo hicieron.

- Con respecto a los destinos:

Las personas que viajaron al resto de América en el pasado, presentan un 1200 % más chances de viajar al *resto del mundo* en un futuro viaje, comparando contra los que no viajaron, si comparamos con limitarse a los países limítrofes.

Las personas que viajaron a Chile, Bolivia o Perú presentan un 300 % más chances de elegir viajar a Europa, comparando con los que no lo hicieron. A su vez, las personas que declararon tener una situación económica muy buena presentan 260 % más chances de elegir Europa como destino, comparando con el resto.

- Con respecto a la modalidad de organización:

Las chances de elegir una agencia por sobre la organización por cuenta propia aumentan 660 % cuando las personas son pudientes.

De acuerdo a los resultados obtenidos, se obtiene una clara segmentación de los prospectos y clientes, y de esta manera, se optimiza la administración y entrega de programas de marketing, puntualmente el *marketing de bases de datos*.

Así pues, ya habiendo obtenido estos modelos, se pudieron desarrollar algunas aplicaciones prácticas que demuestran el valor económico que genera la investigación realizada.

9.3 Sumario de Aplicaciones Prácticas

En una primera etapa, estas aplicaciones consistieron en comparar distintas campañas de promoción de viajes. Para este fin, fue necesario el uso de técnicas estadísticas y de simulación, como comparación de medias con números aleatorios comunes y Simulación de Montecarlo.

Los tres casos fueron los siguientes:

9.3.1 Comparación de Alternativas de envío de Publicidad de Viajes con Hoteles de Lujo para Jóvenes.

En una primera aproximación, se envió a toda una lista de personas, la oferta de un viaje que incluyera un hotel de lujo.

Ya que existía un grupo de personas, al cual no resultaba rentable enviarle publicidad de viajes que incluyera hospedaje de este tipo, se consideró la posibilidad de no enviarles ningún tipo de correspondencia a éstos, surgiendo de ese modo una nueva alternativa, la segunda.

La última y tercer alternativa a comparar, fue la de incurrir en extra costos para realizar más diseños de publicidad, en los que se ofrecieran hospedajes de menor nivel a las personas “poco rentables” de la base, y hospedaje de lujo al resto.

Resumiendo, ya que la última alternativa superaba, en valores esperados de utilidad, en un 31% a la primera y en un 23 % a la segunda, se demostró que era la mejor opción. A su vez, resultó un 16 % más rentable que la primera alternativa.

De esta forma, la mejor estrategia a seguir pareciera ser la de enviar a las personas con probabilidades menores al 50 % de contratar un hotel de lujo, la oferta de un viaje que incluya alojamiento de menor nivel. Al resto de las personas, sí se le envía ofertas que involucren hoteles de categoría.

9.3.2 Comparación de Alternativas de envío de Publicidad de Viajes con Distintos Destinos para Jóvenes.

En una primera aproximación, se envió a toda una lista de personas, la oferta de un viaje que tenga el destino *Resto del Mundo*.

Ya que se demostró que resultaba rentable enviarle publicidad de viajes que incluyan como destino el *resto del mundo* a todos los grupos, la posibilidad de no enviar ningún tipo de publicidad a un grupo determinado, sólo tendría sentido de tener un presupuesto limitado. En ese caso, por ejemplo, se les podría enviar publicidad solamente a las personas que sí han viajado al Resto de América, ya que sería el conjunto más lucrativo, de esta manera nace la segunda alternativa.

Ahora bien, en lugar de no enviar ningún tipo de publicidad al grupo que no ha viajado al resto de América, surge la posibilidad de enviarles ofertas de viajes en la que el destino sea un país limítrofe. De este modo, se incurre en un costo extra ya que se realizarán más diseños, unos son los que fueron utilizados en las alternativas anteriores, y los nuevos en los que se ofrece esta nueva posibilidad, surgiendo de esta forma la tercera alternativa.

Así pues, la mejor estrategia estará ligada a los costos en los que se puede y quiere incurrir. La alternativa de menor costo sería la segunda, que a su vez es la más rentable. De poseer más presupuesto, se podría optar entre las primer y tercer opción. La tercera alternativa mejora las utilidades esperadas en un 6 %

con respecto a la primera, pero los costos asociados de esta posibilidad, aumentan un 14 %.

El tercer caso fue distinto, en esta oportunidad, se hizo uso de un árbol de decisión, herramienta comúnmente utilizada en inversión en opciones reales en el área de finanzas y estrategia corporativa.

9.3.3 Comparación de Alternativas de Publicidad para Distintas Modalidades de Organización de Viajes.

Recordando de los modelos obtenidos en la sección anterior, las modalidades de organización de viajes estudiadas en este trabajo eran:

- por contacto con una Agencia de Viajes
- por Internet
- por Cuenta Propia

La idea en esta oportunidad será diseñar una campaña de publicidad que implique el mayor valor económico posible, en este caso, para una agencia de viajes. Para tal fin, se evaluaron distintas opciones que luego fueron evaluadas en un árbol de decisión.

Se valoraron cinco opciones. La última de éstas a su vez tenía 4 sub-opciones asociadas. Se colocó toda esta información en un árbol de decisión para la correcta decisión de la mejor alternativa.

La mejor opción resultó la que realizaba dos pruebas pilotos y, dependiendo del éxito o fracaso de estos tests, se actuaba de la manera más conveniente. Así se obtuvo el mejor resultado, ya que superaba en un 46 % o más a las utilidades del resto de las alternativas.

Por lo tanto, los pasos a seguir de querer realizar una campaña de este tipo, serían realizar pruebas pilotos y actuar en vista del éxito o fracaso de ellas. De esta forma, se aseguran los mejores valores económicos y estratégicos desde el punto de vista del Marketing.

Vale aclarar que, haciendo a un lado a todo lo referido al turismo, un valor extra de este proyecto reside en la explicación detallada de la obtención de modelos de regresión logística, tanto binomiales como multinomiales. Éstos, son de gran utilidad en la actualidad dada la potencia de las herramientas informáticas, pudiendo resolver problemas que en el pasado hubieran requerido un increíble esfuerzo de cálculo.

Son aplicados en numerosos ámbitos como pueden ser la consultoría para diversas empresas, el desarrollo de correctas segmentaciones en el Marketing, estimación de probabilidades de siniestros u otras ilimitadas aplicaciones en la

ciencia, la investigación y la tecnología. Los límites sólo residen en la capacidad de la mente humana para encontrar donde aplicarlos.

9.4 Límites de la Investigación

A pesar de que las conclusiones que se han obtenido en este trabajo son significativas, no sólo desde el punto de vista estadístico sino también desde lo lógico y racional, es preciso aclarar que la interpretación de los modelos deben realizarse teniendo en cuenta que la investigación ha estado sujeta a ciertas limitaciones, lo que propone futuros trabajos para superarlas.

Se recuerda que, los datos relevados en el trabajo de campo son en muchos casos valoraciones subjetivas de los encuestados, puesto que las preguntas apuntaban a la percepción acerca de diversas variables independientes. Al tratarse de percepciones, se maneja cierta subjetividad que inexorablemente desemboca en desvíos. A pesar de lo enunciado, no se encontraron respuestas ilógicas, por lo que las posibles desviaciones podrían ser simplemente cuestión de entendibles diferencias de apreciación.

Otras limitaciones consisten en la manera de obtención de los datos para crear las bases. En nuestro caso, las respuestas fueron realizadas por gente conocida, que lo hizo de favor. Es de esperar, que eso no suceda con una empresa u organización con fines de lucro. La obtención de una base en este caso no es explicada en esta investigación. Sin embargo, muchas empresas ya poseen bases de datos, o tienen los medios económicos para acceder a ellas. Así pues, por ejemplo, se podrían remplazar los datos económicos subjetivos en nuestro estudio con datos objetivos reales de las personas.

También hay límites en cuanto a la teoría del marketing por base de datos. Hay ciertas cuestiones que no han sido tenidas en cuenta y deben ser mencionadas. Lo primero que se debe aclarar, es que el este tipo de marketing requiere de mucha inversión, ya sea en hardware, software, enlaces de telecomunicaciones y personal capacitado, como en el diseño, recopilación y actualización permanente de las bases de datos.

A todo esto se suma el dilema ético que presupone la venta de datos de personas, ya que entra en juego la intimidad y privacidad de éstos.

9.5 Acerca de Futuras Investigaciones

A partir de los resultados y limitaciones de nuestro estudio, surgen ineluctablemente futuras líneas de investigación, que contribuirían con el enriquecimiento del análisis realizado. Se recuerda que han sido relevadas personas mayoritariamente de clase media alta, residentes en Capital Federal y con una gran proporción de jóvenes. Pues entonces surge intuitivamente la posibilidad de incorporar mayor cantidad de personas, preferentemente de otros orígenes, edades y lugares de residencia, a fin de que los alcances del proyecto sean mayores.

Se recuerda también, que se ha hecho foco en el turismo emisor, el de los turistas residentes en la Argentina que viajan al exterior. Por lo tanto, en un futuro trabajo, se podría estudiar el turismo receptor, de gran auge en la actualidad debido al tipo de cambio favorable para la recepción de extranjeros provenientes mayoritariamente de América y Europa.

Otro punto a ampliar son la cantidad de variables independientes y las categorías de las variables dependientes. De ampliar la cantidad de variables independientes, se pueden obtener segmentaciones mucho más precisas y concluyentes. Con respecto a la mayor cantidad de categorías de variables dependientes, esta posibilitaría un abanico de opciones que aportaría flexibilidad al estudio. Vale aclarar que para que esto sea posible, se necesitarían tamaños muestrales bastante mayores, que de ser financiados y realizados correctamente podrían devenir en grandes beneficios para los interesados en cuestión.

Por último, se podrían relevar en las encuestas, más cuestiones de interés relacionadas con el turismo. Estas podrían ser, duraciones planeadas de los viajes de placer o épocas del año preferidas o disponibles para viajar, entre otras tantas.

BIBLIOGRAFÍA

McCullagh, P., and J. A. Nelder. 1989. Generalized Linear Models, 2nd ed. London: Chapman & Hall. ISBN 0-412-31760-5

Albert, A. and Harris, E. 1987. Multivariate Interpretation of Clinical Laboratory Data. Marcel Dekker, New York, New York. ISBN 0-8247-7735-2

Hosmer, D. W., and S. Lemeshow. 2000. Applied Logistic Regression. New York: John Wiley & Sons. ISBN 9780471356325

Kleinbaum, D. G. 1994. Logistic Regression: A Self-Learning Text. New York: Springer-Verlag. ISBN 0-387-95397-3

Norusis, M. 2004. SPSS 13.0 Statistical Procedures Companion. Upper Saddle-River, N.J.: Prentice Hall, Inc. ISBN 9780131995277

W. David Kelton, Randall P. Sadowski, David T. Sturrock. 2004. Simulation with Arena, Third Edition. Mc Graw Hill. ISBN 0-07-285694-7

Jerry Banks, John S. Carson II, Barry L. Nelson. Discrete-Event Simulation, Second Edition. Prentice Hall, ISBN 0-13-217449-9

Philip Kotler. 2000. Dirección de Marketing, La edición del milenio. Pearson Education. ISBN 968-444-422-3

Roberto M. García. 2004. Inferencia Estadística y diseño de Experimentos. Eudeba. ISBN 950-23-1295-3

Zvi Bodie, Robert C. Merton. Finanzas, Primera edición revisada. Pearson, Prentice Hall. ISBN 970-26-0097-9.

ANEXOS

Anexo I. Encuesta

Nombre (opcional):

Marque con una X sus respuestas a las siguientes preguntas:

1. Sexo:

- Masculino
- Femenino

2. Estado Civil:

- Soltero/a, divorciado/a o viudo
- Casado/a o en pareja
- Casado/a con hijos/as menores a 18 años

3. Edad:

- 15 a 30
- 31 a 45
- 46 a 60
- Más de 60.

4. De haber viajado como turista en los últimos 5 años, con qué destino/s lo hizo? (Marque una o más respuestas)

Destinos:

- Chile, Bolivia o Perú
- Centroamérica
- Uruguay
- Brasil
- Europa
- Norteamérica (EE UU, Canadá)
- Oceanía o Africa
- No viajé
- Argentina (**Indique Costa, Sur , Norte, etc**) u otro (Indique país):

Utilizó hoteles de 3 o más estrellas en esos viajes?

- Si
- No

De haber viajado, cómo planeó su viaje? (Marque una o más respuestas)

- Contacto con agencia de viajes
- Lo planee por Internet
- Otro (Indique cómo):

5. Planea viajar como turista dentro de los próximos 2 años, con qué destino/s lo haría? (Marque una o más respuestas)

Destinos:

- Chile, Bolivia o Perú
- Centroamérica
- Uruguay
- Brasil
- Europa
- Norteamérica (EE UU, Canadá)
- Oceanía o Africa
- No planeo viajar
- Argentina (**Indique Costa, Sur , Norte, etc**) u otro (Indique país):

Utilizaría hoteles de 3 o más estrellas en estos futuros viajes?

- Si
- No

De tener pensado viajar, cómo planearía su viaje? (Marque una o más respuestas)

- Contacto con agencia de viajes
- Lo planearía por Internet
- Otro (Indique cómo):

6. Del dinero disponible para gastar en cosas que le gustan. Cuánto destina a viajes turísticos?

- Poco.
- Medio
- Mucho

7. Cómo considera su situación económica?

- Mala (no podría/no me podrían pagar un pequeño viaje con destino local)

- Regular (Sólo podría/me podrían pagar un viaje con destino local pero pocos días)
- Buena (Podría/me podrían pagar un viaje a un destino local un par de semanas o pagar un viaje al exterior pero no más de una semana)
- Muy Buena (Podría/ me podrían pagar un viaje al exterior de un par de semanas o más)

8. Estudios:

- Secundario
- Universitario incompleto/ en curso
- Universitario completo

9. Trabaja?

- Si
- No

Anexo II. Técnicas estadísticas²⁹

Métodos Factoriales (de interdependencia)

Análisis de Componentes Principales (H. Hotelling)

Variables:

-Cuantitativas.

Objetivo Fundamental:

-Obtener una representación sintética del conjunto de individuos, con un grupo menor de variables, que se establecen como combinaciones lineales entre sí de las variables originales.

Ecuaciones estructurales

Objetivo Fundamental:

-Estudio de las relaciones causales entre variables. Utilizando estos modelos se puede estimar efectos entre variables y contrastar modelos de interdependencia entre ellas.

Análisis cluster

Objetivo Fundamental:

-Agrupar las observaciones de forma que los datos sean muy homogéneos dentro de los grupos (mínima varianza) y que estos grupos sean lo más heterogéneos posible entre ellos (máxima varianza). De este modo se obtiene una clasificación de los datos multivariante con la que podemos comprender mejor los mismos y la población de la que proceden. Se pueden realizar análisis cluster de casos, análisis cluster de variables o análisis cluster por bloques si se agrupan variables y casos.

Análisis Factorial Exploratorio (Ch. Spearman, L. Thurstone)

Variables:

²⁹ Las explicaciones del Análisis de Componentes Principales, el Análisis Factorial Exploratorio, el Análisis Multivariante de la Varianza, el Análisis Multivariante de la Covarianza, el Análisis de Correlación Canónica y el Análisis Discriminante se encuentran en el libro "Inferencia Estadística y Diseño de Experimentos" de Roberto Mariano García, Capítulo 23, a partir de la página 560.

-Cuantitativas.

Objetivo Fundamental:

-Establecer las causales ocultas o factores latentes de la variabilidad observada.

Ejemplo: Se realiza una encuesta a un conjunto de consumidores de un producto sobre la importancia de sus características. Por ejemplo, en una encuesta sobre insecticidas domésticos se pregunta sobre la importancia asignada a: 1) toxicidad; 2) si puede aplicarse en el cuarto de los niños; 3) su efectividad; 4) su poder residual; 5) su olor, agradable o no; 6) si es efectivo para todo tipo de insectos; 7) si su envase es práctico; etcétera. Con los datos de la encuesta, el Análisis Factorial descubrirá los factores latentes, es decir, lo que verdaderamente le interesa a los consumidores y que no es evidente *a priori*, ni siquiera para ellos mismos, pues ese interés radica en su subconsciente. Luego, al conocerse esos factores, se fabricará un producto que lo tenga en cuenta y se diseñará la publicidad adecuada.

Métodos de regresión (de dependencia)

Análisis Multivariante de la Varianza (Manova) (Confirmatorio) (R. Fisher)

Tipo de variables:

-Explicativas: Una o más cualitativas

-Explicadas: Una o más cuantitativas

Objetivo Fundamental:

-Establecer la variabilidad observada en los datos obtenidos de un experimento e imputar dicha variabilidad a las distintas causas o factores de variación.

Análisis Multivariante de la Covarianza (Mancova) (Confirmatorio) (R. Fisher)

Tipo de variables:

-Explicativas: Una o más cualitativas y una o más cuantitativas.

-Explicadas: Una o más cuantitativas.

Objetivo Fundamental:

-Estudiar la variabilidad observada en los datos obtenidos de un experimento e imputar dicha variabilidad a las distintas causas o factores de variación, considerando una o más variables cuantitativas para el control del error experimental.

Análisis de Correlación Canónica (Exploratorio) (H. Hotelling)

Tipo de variables:

-Explicativas (X): Dos o más cuantitativas.

-Explicadas (Y): Dos o más cuantitativas.

Objetivo Fundamental:

-Obtener combinaciones lineales de las variables X [$U = f(X)$] y de las variables Y [$V = g(Y)$], cuyas correlaciones $R(U, V)$ sean máximas.

Análisis Discriminante (Confirmatorio) (R. Fisher)

Tipo de variables:

-Explicativas (X): Una o más cuantitativas.

-Explicada (Y): Una cualitativa (categórica).

Objetivo Fundamental:

-Clasificar nuevos individuos en alguna de las categorías (grupos o *clusters*) de la variable explicada, en función de los valores que toman las variables explicativas (o discriminantes) en esos individuos. Se lo puede considerar también un método factorial.

Regresión Logística³⁰

La regresión logística estudia la asociación entre una variable dependiente categórica y un juego de variables independientes (explicativas). El nombre *regresión logística* es comúnmente utilizado cuando la variable dependiente tiene

³⁰ Traducido y adaptado de NCSS User's guide III, 2006. Chapter 320.

sólo 2 valores (aunque técnicamente se trate de Regresión Logística Binaria). Regresión logística multinomial es el nombre que se reserva cuando la variable dependiente tiene 3 o más valores únicos.

La regresión logística multinomial también se conoce como regresión logística politómica.

La regresión logística compete con el análisis discriminante como un método para analizar variables de respuesta discretas. Es más, el sentimiento entre las personas que se dedican a la estadística como profesión es que la regresión logística es más versátil y más apta para más situaciones que el análisis discriminante porque no asume que las variables independientes están normalmente distribuidas como el caso del análisis discriminante.

Las Transformaciones Logit y Logística

En regresión múltiple, un modelo matemático de un juego de variables explicativas se usa para predecir la media de una variable dependiente. En regresión logística, un modelo matemático de un juego de variables explicativas se usa para predecir una transformación de la variable dependiente. A esto se le llama transformación logit.

Supongamos que los valores numéricos 0 y 1 son asignados a las dos categorías de una variable binaria. Frecuentemente, el 0 representa una respuesta negativa y 1 representa una respuesta positiva. La media de esta variable será la proporción de respuestas positivas. Debido a esto, se podría tratar de modelar la relación entre la probabilidad (proporción) de una respuesta positiva y las variables explicativas.

Si p es la proporción de observaciones con respuesta 1, entonces $1-p$ es la probabilidad de una respuesta 0. El ratio $p/(1-p)$ es llamado odds (en español serían las chances) y logit es el logaritmo natural de las chances, llamado también log odds. Matemáticamente la transformación logit se escribe:

$$l = \text{logit}(p) = \ln\left(\frac{p}{1-p}\right)$$

La siguiente tabla muestra el logit para varios valores de p :

P	Logit(P)
0,001	-6,91
0,01	-4,60
0,05	-2,94
0,5	0,00
0,8	1,39
0,999	6,91

Nótese que mientras el rango de p va de cero a uno, el rango logit va desde menos a más infinito. Además se muestra que el cero logit ocurre cuando p es 0,5.

La transformación logística es la inversa de la transformación logit. Se escribe:

$$p = \text{logistic}(l) = \frac{e^l}{1 + e^l}$$

La transformación “Log Odds Ratio”

La diferencia entre dos log odds puede ser utilizada para comparar dos proporciones, como por ejemplo hombres versus mujeres. Matemáticamente, la diferencia es escrita:

$$\begin{aligned} l_1 - l_2 &= \text{logit}(p_1) - \text{logit}(p_2) = \ln\left(\frac{p_1}{1-p_1}\right) - \ln\left(\frac{p_2}{1-p_2}\right) = \\ &= \ln\left(\frac{\left(\frac{p_1}{1-p_1}\right)}{\frac{p_2}{1-p_2}}\right) = \ln\left(\frac{p_1(1-p_2)}{p_2(1-p_1)}\right) = \ln(OR_{1,2}) \end{aligned}$$

Esta diferencia es usualmente conocida como log odds ratio. El odds ratio es frecuentemente utilizado para comparar proporciones entre grupos. Nótese que la transformación logística es cercanamente relacionada a los odds ratios. La relación inversa es:

$$OR_{1,2} = e^{(l_1 - l_2)}$$

La Regresión logística y los modelos Logit

En regresiones logísticas multi-grupo, a una variable discreta dependiente Y que tiene G únicos valores ($G \geq 2$) se le realiza la regresión con un juego de p variables independientes X_1, X_2, \dots, X_p . Y representa una manera de particionar la población de interés. Por ejemplo, Y puede ser la ausencia o presencia de una

enfermedad, condición luego de una cirugía, o estado civil. Como el nombre de las particiones es arbitrario, nos referiremos a ellas con números consecutivos. Por lo tanto, Y tomará los valores 1, 2, ..., G .

En las explicaciones siguientes utilizaremos:

$$X = (X_1, X_2, \dots, X_p)$$

$$B = \begin{pmatrix} \beta_{g1} \\ \vdots \\ \beta_{gp} \end{pmatrix}$$

El modelo de regresión logística es dado por las G ecuaciones:

$$\begin{aligned} \ln\left(\frac{p_g}{p_1}\right) &= \ln\left(\frac{P_g}{P_1}\right) + \beta_{g1} X_1 + \beta_{g2} X_2 + \dots + \beta_{gp} X_p = \\ &= \ln\left(\frac{P_g}{P_1}\right) + XB_g \end{aligned}$$

Aquí, p_g es la probabilidad de que un individuo con valores X_1, X_2, \dots, X_p esté en el grupo g . Esto es:

$$p_g = Pr(Y = g | X)$$

Usualmente $X_1 = 1$ (ya que se incluye un intercepto), pero esto no tiene por qué ser necesariamente así. Las cantidades P_1, P_2, \dots, P_g representan las probabilidades a priori de pertenecer a un grupo. Si estas probabilidades a priori se asumen que son iguales, entonces el término $\ln(P_g/P_1)$ se hace cero y desaparece. Si a priori no son iguales, cambian los valores de los interceptos en la ecuación de regresión logística.

Al grupo uno se lo llama *grupo de referencia*. Los coeficientes de regresión $\beta_{11}, \beta_{12}, \dots, \beta_{1p}$ para el grupo de referencia son fijados como cero. La elección del grupo de referencia es arbitraria. Usualmente, es el grupo más grande o un grupo de control el cual quiere ser comparado con otros grupos. Esto deja $G-1$ ecuaciones de regresión logística en el modelo logística multinomial.

Los β 's son coeficientes de regresión poblacionales que tendrán que ser estimados a partir de los datos. Sus estimados son representados por los b 's. Entonces, los β 's representan los parámetros desconocidos, mientras que los b 's son sus estimados.

Estas ecuaciones son lineales en los logits de p . Sin embargo, en términos de probabilidades, son no lineales. Las correspondientes ecuaciones no lineales son:

$$p_g = \text{Prob}(Y = g|X) = \frac{e^{XB_g}}{1 + e^{XB_2} + e^{XB_3} + \dots + e^{XB_G}}$$

ya que $e^{XB_1} = 1$ porque todos los coeficientes de esta regresión son cero.

Una nota acerca de la interpretación de e^{XB} puede ser útil, utilizando el hecho de que $e^{a+b} = (e^a)(e^b)$, e^{XB} puede ser representado como sigue:

$$e^{XB} = e^{\beta_1 X_1 + \beta_2 X_2 + \dots + \beta_p X_p}$$

Esto muestra que el valor final es el producto de los términos individuales.

El caso particular de la Regresión Logística Binaria

Este caso particular de Regresión Logística resulta el más conocido y utilizado en muchos ámbitos. Se utiliza cuando se quiere estimar la probabilidad de ocurrencia de un evento para una variable de respuesta categórica que sólo puede tomar dos valores.

Por ejemplo, se puede dar el caso de una entidad prestamista que quiere estimar si un nuevo tomador de un préstamo es probable que sea moroso y no pague. Con esta herramienta el prestamista puede evaluar el riesgo de entregarle un préstamo a determinada persona.

Ya que la probabilidad de un evento debe estar entre 0 y 1, resulta impráctico modelar probabilidades con técnicas de Regresión Lineal, porque estos modelos permiten que la variable dependiente tome valores tanto mayores a 1 como menores a 0. El modelo de Regresión Logística Binaria es un tipo de modelo lineal

generalizado que extiende el modelo de Regresión Lineal General vinculando el rango de números reales al de 0 y 1.

Se empieza por considerar la existencia de una variable continua inadvertida, Z , que puede ser pensada como la “propensión hacia” el evento de interés. En el caso del prestamista, Z representa la propensión de un cliente a ser moroso en un préstamo, con grandes valores de Z que se corresponden con altas probabilidades de ser moroso.

En el modelo, la relación entre Z y la probabilidad del evento de interés es descrita mediante la siguiente función vinculante:

$$\pi_i = \frac{e^{z_i}}{1 + e^{z_i}} = \frac{1}{1 + e^{-z_i}}$$

Ó

$$z_i = \log\left(\frac{\pi_i}{1 - \pi_i}\right)$$

donde:

π_i es la probabilidad de que el caso i ésimo experimente el evento de interés.

z_i es el valor de la variable continua inadvertida para el caso i ésimo.

El modelo también asume que Z está linealmente vinculada con los predictores.

$$z_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_p x_{ip}$$

donde:

x_{ij} es la variable predictora j ésima para el caso i ésimo.

β_j es el coeficiente j ésimo para esa variable.

p es el número de variable predictoras.

Si Z fuera no fuera una variable inadvertida, simplemente se tendría que ajustar una regresión lineal a Z . Como esto no ocurre, se deben relacionar las variables predictoras a la probabilidad de interés mediante la sustitución de Z .

$$\pi_i = \frac{1}{1 + e^{-(b_0 + b_1 x_1 + \dots + b_p x_p)}}$$

Los coeficientes de Regresión se estiman a través de un método iterativo de Máxima Verosimilitud.

Resolviendo las Ecuaciones de Verosimilitud

Para mejorar la notación, se escribe:

$$\begin{aligned} \pi_{g|} = \text{Prob}(Y = g|X) &= \frac{e^{XB_g}}{1 + e^{XB_2} + e^{XB_3} + \dots + e^{XB_G}} = \\ &= \frac{e^{XB_g}}{G} \\ &= \frac{e^{XB_g}}{\sum_{s=1}^G e^{XB_s}} \end{aligned}$$

La similitud para una muestra de N observaciones viene dada por:

$$l = \prod_{j=1}^N \prod_{g=1}^G \pi_{g|}^{y_{gj}}$$

donde y_{gj} es uno si la observación j es en el grupo g y cero de otro modo.

Utilizando el hecho de que:

$$\sum_{s=1}^G y_{gs} = 1$$

la similitud log, L , es dada por:

$$L = \ln(l) =$$

$$= \sum_{j=1}^N \left[\sum_{g=1}^G y_{gj} X_j B_g - \ln \left(\sum_{g=1}^G e^{X_j B_g} \right) \right]$$

Los estimados de máxima verosimilitud de los β 's se encuentran buscando los valores que maximicen la ecuación de log similitud. Esto se consigue calculando las derivadas parciales e igualándolas a cero. Las ecuaciones de similitud resultantes son:

$$\frac{\partial L}{\partial \beta_{kg}} = \sum_{j=1}^N x_{kj} (y_{jg} - \pi_{jg})$$

para $g = 1, 2, \dots, G$ y $k = 1, 2, \dots, p$. En realidad, como todos los coeficientes son cero para $g = 1$, el rango de g es de 2 a G .

Debido a la naturaleza no lineal de los parámetros, no hay solución cerrada a estas ecuaciones y éstas deben ser resueltas por iteración. El método de Newton-Raphson como se describe en Albert and Harris (1987) se utiliza para resolver las ecuaciones. Este método hace uso de la matriz de información, $I(\beta)$, el cual es formado a partir de la matriz de las segundas derivadas parciales. Los elementos de la matriz de información están dados por:

$$\begin{aligned} \frac{\partial^2 L}{\partial \beta_{ik} \partial \beta_{i'k'}} &= - \sum_{j=1}^N x_{kj} x_{k'j} \pi_{ij} (1 - \pi_{ig}) = \\ &= \frac{\partial^2 L}{\partial \beta_{ik} \partial \beta_{i'k'}} = \sum_{j=1}^N x_{kj} x_{k'j} \pi_{ig} \pi_{i'g} \end{aligned}$$

La matriz de información es utilizada porque la matriz de covarianza asintótica de los estimados de máxima verosimilitud es igual a la inversa de la matriz de información. Esto es:

$$V \begin{pmatrix} \hat{\Lambda} \\ \hat{\beta} \end{pmatrix} = I(\hat{\beta})^{-1}$$

Esta matriz de covarianza es usada en los cálculos de intervalos de confianza para los coeficientes de regresión, odds ratios, y probabilidades predichas.

Interpretación de los coeficientes de regresión

La interpretación de los coeficientes estimados de la regresión no es tan fácil como en regresión múltiple. En regresión logística multinomial, no sólo la relación entre X e Y es no lineal, sino que además si la variable dependiente tiene más de dos valores únicos, las ecuaciones de regresión son varias.

Considerando el simple caso de una variable dependiente binaria, Y, y una sólo variable independiente, X. Asumamos que Y es codificada de tal modo que sólo tome los valores 0 o 1. En este caso, la ecuación de regresión logística es:

$$\ln \left(\frac{p}{1-p} \right) = \beta_0 + \beta_1 X$$

Ahora consideremos un incremento de una unidad en la variable X. La ecuación de regresión logística resulta:

$$\ln\left(\frac{p'}{1-p'}\right) = \beta_0 + \beta_1(X+1) = \beta_0 + \beta_1 X + \beta_1$$

Podemos aislar la pendiente tomando la diferencia de estas dos ecuaciones. Tenemos entonces:

$$\begin{aligned}\beta_1 &= \beta_0 + \beta_1(X+1) - \beta_0 + \beta_1 X = \\ &= \ln\left(\frac{p'}{1-p'}\right) - \ln\left(\frac{p}{1-p}\right) = \\ &= \ln\left(\frac{\left(\frac{p'}{1-p'}\right)}{\frac{p}{1-p}}\right) = \ln\left(\frac{\text{odds}'}{\text{odds}}\right)\end{aligned}$$

Esto es, β_1 es el logaritmo natural del ratio de las chances de $X+1$ y X . Si quitamos el logaritmo natural exponenciando ambos lados resulta:

$$e^{\beta_1} = \frac{\text{odds}'}{\text{odds}}$$

El coeficiente de regresión β_1 es interpretado como el logaritmo natural del ratio de chances comparando las chances después del incremento en una unidad de la variable X con respecto a las chances originales. Nótese que, al contrario de la regresión múltiple, la interpretación de β_1 depende del valor particular de X ya que los valores de probabilidad, los p 's, varían para diferentes X .

X Binaria

Cuando X puede tomar sólo dos valores, digamos 0 y 1, la interpretación de arriba se hace más simple. Como sólo hay dos posibles valores de X , hay una única interpretación para β_1 dada por el logaritmo natural del ratio de las chances. En términos matemáticos, el significado de β_1 es:

$$\beta_1 = \ln \left(\frac{\text{odds}(X = 1)}{\text{odds}(X = 0)} \right)$$

Para entender más esta ecuación, consideremos primero qué son las chances (odds). Las chances son el ratio de dos probabilidades, p y $1-p$. Consideremos la siguiente tabla con valores de chances para varios valores de p . Nótese que 9:1, se lee como “9 a 1”.

<u>valor de p</u>	<u>chances (odds) de p</u>
0,9	9:1
0,8	4:1
0,6	1,5:1
0,5	1:1
0,4	0,67:1
0,2	0,25:1
0,1	0,11:1

Ahora, utilizando un simple ejemplo de carreras de caballos, si un caballo tiene 8:1 chances de ganar y un segundo caballo tiene 4:1 de chances de ganar, ¿Cómo se compara a estos 2 caballos? Una manera obvia es mirar el ratio de sus chances. El primer caballo tiene el doble de chances de ganar que el primero.

Consideremos un segundo ejemplo de dos caballos lentos cuyas chances de ganar son 0,1:1 y 0,05:1. Aquí de nuevo, su ratio de chances es 2. El mensaje es el siguiente: El ratio de chances da un número relativo. Aunque el primer caballo tiene el doble de chances de ganar la carrera es probable que esto no suceda.

Para entender mejor β_1 , debemos tomar el logaritmo natural del ratio de chances. Es difícil pensar en términos de logaritmos, sin embargo, se recuerda que el logaritmo natural de uno es cero. Entonces, un valor positivo de β_1 indica que las chances del numerador son más grandes mientras que un valor negativo de β_1 indica que las chances del denominador son más grandes.

En términos de probabilidad, es más fácil pensar en términos de e^{β_1} que de β_1 , porque e^{β_1} es el ratio de las chances mientras que β_1 es el logaritmo natural del ratio de las chances.

Múltiples Variables Independientes

Cuando hay múltiples variables independientes, la interpretación de cada coeficiente de regresión se hace más difícil. Pero como regla general, la interpretación de arriba es correcta.

Variable Dependiente Multinomial

Cuando la variable dependiente tiene más de dos valores, va a haber más de una ecuación de regresión. De hecho, el número de ecuaciones de regresión es igual a uno menos que el número de valores. Esto hace la interpretación más difícil porque hay varios coeficientes de regresión asociados a cada variable independiente. En este caso, hay que tener cuidado para comprender qué es lo que cada ecuación está prediciendo. Una vez que esto es comprendido, la interpretación de cada uno de los $K-1$ coeficientes de regresión para cada variable puede proceder como se muestra a continuación:

Consideremos el siguiente ejemplo en el cuál hay dos variables independientes, X_1 y X_2 , y la variable dependiente tiene tres grupos: A, B y C.

<u>Fila</u>	<u>Y</u>	<u>X1</u>	<u>X2</u>	<u>GA</u>	<u>GB</u>	<u>GC</u>
1	A	3,2	5,8	1	0	0
2	A	4,7	6,1	1	0	0
3	B	2,8	3,5	0	1	0
4	B	3,3	4,6	0	1	0
5	B	3,9	5,2	0	1	0
6	C	4,2	3,7	0	0	1
7	C	7,3	4,4	0	0	1
8	C	5,3	5,1	0	0	1
9	C	6,8	4,5	0	0	1

Miremos a las tres variables indicativas: GA , GB , y GC . Toman el valor cero dependiendo si Y toma o no el valor correspondiente. Dos ecuaciones de regresión van a ser generadas correspondiendo a dos de las variables indicativas. El valor que no es utilizado es llamado *valor de referencia*. Supongamos que el valor de referencia es C. Las dos ecuaciones de regresión son:

$$\ln\left(\frac{p_A}{p_C}\right) = \beta_{A0} + \beta_{A1}X_1 + \beta_{A2}X_2$$

y

$$\ln\left(\frac{p_B}{p_C}\right) = \beta_{B0} + \beta_{B1}X_1 + \beta_{B2}X_2$$

Los dos coeficientes para X_1 en estas ecuaciones, β_{A1} y β_{B1} , dan el cambio en el logaritmo natural de A versus C y de B versus C por un cambio en una unidad de X_1 , respectivamente.

Tests Estadísticos e Intervalos de Confianza

Inferencias acerca de coeficientes de regresión, grupos de regresión de coeficientes, bondades de ajuste, respuestas medias y predicciones de membresía de grupo de nuevas observaciones son todos de interés. Estos procedimientos de inferencia pueden ser tratados de considerar tests de hipótesis y/o intervalos de confianza. Los procedimientos de inferencia en regresión logística se fían en grandes tamaños de muestra para adquirir precisión.

Dos procedimientos están disponibles para evaluar la significancia de una o más variables independientes en regresión logística: Razón de Verosimilitud y Tests de Wald. Estudios de simulación usualmente muestran que el test de razón de verosimilitud tiene un mejor resultado que el Test de Wald. Sin embargo, el test de Wald se sigue utilizando para evaluar la significancia de coeficientes de regresión individuales debido a su facilidad de cálculo.

Los dos procedimientos de evaluación serán explicados a continuación:

Razón de Verosimilitud y Desviación

El estadístico Razón de verosimilitud es -2 veces la diferencia entre los logaritmos naturales de las similitudes de dos modelos, uno de los cuales es un subconjunto de otro. La distribución del estadístico LR (del inglés *Likelihood Ratio*) es cercanamente aproximado por la distribución Chi-Cuadrado para grandes tamaños de muestra. Los grados de libertad de la distribución Chi-Cuadrado aproximada son iguales a la diferencia en el número de coeficientes de regresión de los dos modelos. El test es llamado como un ratio, en vez de una diferencia, ya que la diferencia entre dos logaritmos naturales de similitudes es igual al logaritmo natural del ratio de dos similitudes. Esto es, si L_{full} es el logaritmo natural de la similitud del modelo completo (en inglés full model) y L_{subset} (del inglés subconjunto) es el logaritmo natural de la similitud de un subconjunto del modelo completo, la razón de similitud es definida como:

$$LR = -2[L_{subset} - L_{full}] = -2 \left[\ln \left(\frac{l_{subset}}{l_{full}} \right) \right]$$

Nótese que el -2 ajusta LR para que la distribución Chi-Cuadrado pueda aproximar su distribución.

El test de razón de verosimilitud es el test de elección en regresión logística. Varios estudios de simulación han mostrado que es más preciso que el test de Wald en situaciones de pequeños a moderados tamaños de muestra. En grandes muestras, tienen parecidos resultados. Desafortunadamente, el test de razón de verosimilitud requiere más cálculo que el test de Wald ya que necesita que dos modelos de máxima verosimilitud sean resueltos.

Desviación

Cuando el modelo completo en el estadístico del test de razón de verosimilitud es el modelo saturado, LR es referido como la *desviación*. Un modelo saturado es uno que incluya todos los posibles términos (incluyendo interacciones) tal que los valores predichos del modelo igualen los datos originales. La fórmula para la desviación es:

$$D = -2[L_{reduced} - L_{saturated}]$$

La desviación puede ser calculada directamente utilizando la fórmula para la desviación de residuos (discutido abajo). La fórmula es:

$$D = 2 \sum_{j=1}^J \sum_{g=1}^G w_{gj} \ln \left(\frac{w_{gj}}{n_j p_{gj}} \right)$$

Esta expresión puede ser utilizada para calcular el logaritmo natural de la verosimilitud del modelo saturado sin realmente resolverlo. La fórmula es:

$$L_{saturated} = L_{reduced} + D/2$$

La desviación en regresión logística es análoga a la suma de residuos de cuadrados en regresión múltiple. En realidad, cuando la desviación es calculada en regresión múltiple, es igual a la suma de cuadrados residuales. Las desviaciones residuales, pueden ser elevadas al cuadrado y sumadas como un camino alternativo para calcular la desviación, D .

El cambio en la desviación, ΔD , debido a la inclusión (o exclusión) de una o más variables es utilizada en regresión logística igual a los test parciales F utilizados en regresión múltiple. Muchos textos utilizan la letra G para representar ΔD , pero ya hemos utilizado G para representar el número de grupos en Y . En vez de utilizar la distribución F , la distribución del cambio en la desviación es aproximada por la distribución Chi-Cuadrado. Nótese que como el logaritmo natural de la verosimilitud para el modelo saturado es común para ambas desviaciones, ΔD es calculado sin realmente estimar el modelo saturado. Este hecho se torna muy importante durante la elección de subconjuntos. La fórmula para ΔD para evaluar la significancia de el/los coeficiente(s) de regresión asociados con la variable independiente X_1 es:

$$\begin{aligned} \Delta D_{X1} &= D_{sinX1} - D_{conX1} = \\ &= -2[L_{sinX1} - L_{saturada}] + 2[L_{conX1} - L_{saturada}] = \\ &= -2[L_{sinX1} - L_{conX1}] \end{aligned}$$

Nótese que esta fórmula se ve igual a la del estadístico de razón de verosimilitud. Por lo similar entre el cambio en el test de desviación y en el test de razón de verosimilitud, sus nombres son usualmente utilizados intercambiadamente.

Test de Wald

El test de Wald resultará familiar para aquellos que utilizan regresión múltiple. En regresión múltiple, el conocido t-test para evaluar la significancia de un coeficiente de regresión particular es un test de Wald. En regresión logística, el test de Wald es calculado de la misma forma. La fórmula para el estadístico de Wald es:

$$z_j = \frac{b_j}{s_{b_j}}$$

donde s_{b_j} es un estimado del error estándar de b_j provisto por la raíz cuadrada del correspondiente elemento de la diagonal de la matriz de covarianza, $V(\beta)$. Con tamaños de muestra grandes, la distribución de z_j es cercanamente aproximada con la distribución normal. Con pequeños a moderados tamaños de muestra, la aproximación normal es descripta como “adecuada”. El test de Wald es utilizado para evaluar la significatividad estadística de los coeficientes de regresión individuales.

Intervalos de Confianza

Los intervalos de confianza para los coeficientes de regresión están basados en las estadísticas de Wald. La fórmula para los límites de un $100(1-\alpha)\%$ intervalo de confianza bilateral es:

$$b_j = \pm |z_{\alpha/2}| s_{b_j}$$

R-cuadrado

Hosmer y Lemeshow (1989) se manifestaron en contra del uso de R^2_L (el R-Cuadrado Logístico) como medida de bondad de ajuste. Sin embargo, lo incluiremos porque provee una medida comparativa de la proporción del logaritmo natural que es tomado en cuenta por el modelo. Sólo recordemos que un valor de R^2_L igual a 1 indica que el modelo de regresión logística logra el mismo logaritmo natural de la verosimilitud que el modelo saturado. Sin embargo, esto no significa que ajusta los datos perfectamente. En cambio, significa que ajusta los datos tan bien como se podría desear.

Diagnóstico de Residuos

Los residuos son las discrepancias entre los valores de los datos y de sus valores predichos del modelo resuelto. Un análisis de residuos detecta *outliers* (una observación que numéricamente esté distante del resto de los datos), identifica observaciones de influencia, y diagnostica lo apropiado de un modelo logístico. Un análisis de residuos debería ser conducido antes de utilizar un modelo de regresión.

Residuos Simples

Cada una de las g ecuaciones de regresión logística puede ser utilizada para estimar las probabilidades de que una observación de una variable independiente dada por X_j pertenezca a un grupo correspondiente. Los valores reales de estas probabilidades fueron definidas antes como:

$$\pi_{gj} = \text{Prob} (Y = g | X)$$

Los valores estimados de estas probabilidades son llamadas p_{gj} . Si el símbolo del sombrero es utilizado para representar un parámetro estimado, entonces:

$$\hat{p}_{gj} = \pi_{gj}$$

Estas probabilidades estimadas pueden ser comparadas con las probabilidades reales ocurriendo en la base de datos sustrayendo las dos cantidades, formando un residuo. Los valores reales fueron definidos como las variables indicativas y_{gj} . Por lo tanto, los residuos simples pueden ser definidos como:

$$r_{gj} = y_{gj} - \hat{p}_{gj}$$

Nótese que, a diferencia de la regresión múltiple, hay g residuos para cada observación en vez de sólo uno. Esto hace el análisis de residuos mucho más difícil. Si el modelo de regresión logística se adapta a una observación cercanamente, todos sus residuos serán pequeños. Por lo tanto, cuando y_{gj} es uno, p_{gj} va a ser cercano a uno y cuando y_{gj} es cero p_{gj} va a ser cercano a cero.

Desafortunadamente, los residuos simples tienen desigual varianza igual a $n_j \pi_{gj} (1 - \pi_{gj})$ donde n_j es el número de observaciones con los mismos valores de las variables independientes como observación j . Esta varianza desigual hace que las comparaciones entre los residuos simples sean dificultosas y tipos alternativos de residuos son necesarios.

Residuos de Pearson

Una alternativa popular a los residuos simples (y la que utilizaremos en este trabajo) son los residuos de Pearson que son así nombrados porque dan la contribución de cada observación al estadístico de bondad de ajuste Pearson Chi-Cuadrado. Cuando algunas observaciones son duplicadas la fórmula es:

$$\chi_j = \pm \sqrt{\sum_{g=1}^G \frac{(w_{gj} - n_j p_{gj})^2}{n_j p_{gj}}}, \quad j = 1, 2, \dots, J$$

donde el signo más (menos) es utilizado si w_{gj} / n_j es mayor (menor) que p_{gj} .

Por definición, la suma de cuadrados de residuos de Pearson es el estadístico de bondad de ajuste Chi-Cuadrado de Pearson. Esto es:

$$\chi^2 = \sum_{j=1}^J \chi_j^2$$