

Universidad de la República
Facultad de Ciencias Económicas y de Administración
Instituto de Estadística

**Estimación del número de turistas que ingresan
a Uruguay y caracterización de su perfil para los
ingresados en el año 2008.**

Pasantía para la obtención del título de Licenciado en Estadística

Autoras: Florencia Bacigalupi
Ana Rosenbaum

Tutores: Ramón Álvarez
Silvia Rodríguez
Agosto 2010

AGRADECIMIENTOS

Se agradece al Ministerio de Turismo y Deporte del Uruguay por habernos cedido la base de datos proporcionada por la División Investigación y Estadísticas, en particular a Benjamín Liberoff, Víctor Sosa y Francisco Gardil.

También agradecemos especialmente a nuestros tutores, Ramón Álvarez y Silvia Rodríguez por su dedicación y ayuda en la realización de este trabajo.

Este trabajo es dedicado a Martha Arricar de Aldunate y a Sofía Lejzerson Rosenbaum, nuestras fuentes de inspiración.

RESUMEN EJECUTIVO

El turismo es uno de los factores que contribuye a la generación del producto de una economía nacional. Se puede señalar que mientras en 1950 el turismo internacional recibía U\$S 2,1 mil millones, en 2004 había crecido a U\$S 622,7 mil millones. Debido a la creciente importancia del turismo es fundamental que los agentes tanto del sector público como del privado en Uruguay sean capaces de prepararse para recibir a los turistas de la mejor manera.

En este sentido, el presente trabajo plantea dos objetivos que ayuden a los actores del sector turístico a conocer en profundidad aspectos básicos que definen a los turistas que recibe el país. El primer objetivo fue estimar el número de turistas que ingresarán a Uruguay y el segundo, segmentarlos en distintos perfiles.

Para el primer objetivo se utilizaron datos provistos por el Ministerio de Turismo y Deporte del Uruguay. Se contó con tres series temporales de ingreso de turistas, con frecuencias trimestral y mensual y se generó una tercera serie trimestral como la diferencia algebraica de las dos primeras. Para lograr el objetivo planteado se estimaron tres modelos SARIMA-IA (modelo estacional, autorregresivo y de medias móviles con análisis de intervención). Los tres modelos se identificaron de la siguiente manera:

1. Modelo 1 – Ingreso trimestral de turistas a Uruguay (1993.I-2009.I)
2. Modelo 2 - Ingreso mensual de turistas a Uruguay (1996.01 – 2009.07)
3. Modelo 3 – Diferencia algebraica de los dos modelos anteriores con frecuencia trimestral (1996.I-2009.I)

Se advirtió que la crisis del año 2002 tuvo efecto transitorio para los modelos 1 y 2 mientras que para el modelo 3 tuvo efecto permanente. Otra discrepancia que se encontró es que la variable indicatriz Turismo incidió en el ingreso de turistas a Uruguay en los modelos 1 y 2 pero no afectó particularmente al ingreso de turistas del modelo 3.

El estudio permitió concluir que si se considera el modelo 2 las predicciones indicaron que el número de turistas que ingresará a Uruguay en los próximos doce meses disminuirá (6,42%), mientras que si se considera el modelo 1 las predicciones

indicaron que el número de turistas que ingresará a Uruguay en los próximos cuatro trimestres aumentará (6,24%).

Si se considera el modelo 3 es de esperar que el número de turistas que ingresará a Uruguay en los próximos cuatro trimestres aumente ligeramente (0,1%).

Para el segundo objetivo se contó con datos de la Encuesta de Turismo Receptivo 2008 relevados por el Ministerio de Turismo y Deporte de Uruguay. Se plantearon dos análisis, uno con el objetivo de encontrar tipologías de los turistas que ingresaron a Uruguay durante el año 2008 y otro con el objetivo de analizar la evolución temporal de los puntos de salida en los que son encuestados los grupos de viaje a lo largo de los cuatro trimestres del 2008.

En esta parte, los resultados mostraron que en el primer análisis se encontraron cinco grupos y en el segundo análisis se encontraron tres grupos que permitieron definir lineamientos a seguir de acuerdo al perfil del turista que se obtuvo en cada tipología.

ÍNDICE GENERAL

Resumen Ejecutivo	4
1. Introducción	14
2. Objetivos	16
3. Antecedentes	18
4. Resultados	22
4.1. Modelización de la serie trimestral 1993.I – 2009.I	22
4.1.1. Datos	22
4.1.2. Análisis descriptivo de la serie	22
4.1.3. Ajuste del modelo	26
4.1.4. Modelo seleccionado	27
4.1.5. Interpretación del modelo seleccionado	29
4.1.6. Predicción	30
4.1.6.1. Predicción dentro de la muestra	30
4.1.6.2. Predicción fuera de la muestra	31
4.2. Modelización de la serie mensual 1996.01 – 2009.07	33
4.2.1. Datos	33
4.2.2. Análisis descriptivo de la serie	33
4.2.3. Ajuste del modelo	37
4.2.4. Modelo seleccionado	38
4.2.5. Interpretación del modelo seleccionado	40
4.2.6. Predicción	42
4.2.6.1. Predicción dentro de la muestra	42
4.2.6.2. Predicción fuera de la muestra	45
4.3. Comparación de los modelos 1 y 2	46
4.4. Modelización de la diferencia de la serie trimestral 1993.I – 2009.I y la serie mensual 1996.01 – 2009.07	47
4.4.1. Datos	47
4.4.2. Análisis descriptivo de la serie	47
4.4.3. Ajuste del modelo	51
4.4.4. Modelo seleccionado	53
4.4.5. Interpretación del modelo seleccionado	55
4.4.6. Predicción	57
4.4.6.1. Predicción dentro de la muestra	57
4.4.6.2. Predicción fuera de la muestra	58

4.5. Análisis de la tipología de turistas	60
4.5.1. Datos	60
4.5.2. Objetivos	60
4.5.3. Análisis de los datos sin agregar	64
4.5.3.1. Análisis de Componentes Principales	64
4.5.3.2. Análisis de Clusters	68
4.5.4. Análisis de los datos agregados	73
4.5.4.1. Análisis de Componentes Principales	73
4.5.4.2. Análisis de Clusters	82
4.5.4.3. Evolución Temporal de los Puntos de Salida por Trimestre	88
5. Conclusiones y Recomendaciones	93
Bibliografía	100
A. Apéndice Metodológico	102
A.1. Metodología aplicada a las series de tiempo univariadas	102
A.1.1. Modelización SARIMA(p,d,q)(P,D,Q)	102
A.1.2. Supuestos del modelo univariado	103
A.1.3. Metodología aplicada para la identificación del modelo	104
A.1.4. Análisis de Intervención	113
A.2. Metodología multivariada aplicada a los microdatos de la Encuesta de Turismo Receptivo 2008	117
A.2.1. Análisis de Componentes Principales (ACP)	117
A.2.2. Análisis de Clusters	120
B. Apéndice de Datos	125
B.1. Ingreso trimestral de turistas a Uruguay (1993.01 – 2009.01)	125
B.2. Ingreso mensual de turistas a Uruguay (1996.01 – 2009.07)	126
C. Apéndice de Resultados del Modelo 1	128
C.1. Cumplimiento de los supuestos de los residuos para el modelo seleccionado: ARIMA(1,0,0)(0,1,0)(4) con Turismo y Tc2002	128
C.2. Modelo descartado: ARIMA(1,0,0)(0,1,0)(4) con Turismo y Ls2002	129
C.3. Comparación de modelos	129
C.4. Predicción dentro de la muestra actualizando a un paso para el modelo seleccionado: ARIMA(1,0,0)(0,1,0)(4) con Turismo y Tc2002	130
D. Apéndice de Resultados del Modelo 2	131
D.1. Cumplimiento de los supuestos de los residuos para el modelo seleccionado:	

ARIMA (1,0,0)(0,1,1)(12) con Turismo, Tc Enero 2002 y Tc Mayo 2002	131
D.2. Modelo descartado: ARIMA(1,0,0)(1,1,0)(12) con Turismo, Tc Enero 2002 y Tc Mayo 2002	132
D.3. Comparación de modelos	132
D.4. Predicción dentro de la muestra actualizando a un paso para el modelo seleccionado: ARIMA(1,0,0)(0,1,1)(12) con Turismo, Tc Enero 2002 y Tc Mayo 2002	133
E. Apéndice de Resultados del Modelo 3	136
E.1. Cumplimiento de los supuestos de los residuos para el modelo seleccionado: ARIMA(0,0,1)(0,1,1)(4) con Ls2002	137
E.2. Modelo descartado: ARIMA(1,0,0)(0,1,1)(4) con Ls2002	137
E.3. Predicción dentro de la muestra actualizando a un paso para el modelo seleccionado: ARIMA(0,0,1)(0,1,1)(4) con Ls2002	138
F: Apéndice de Resultados de Tipología de Turistas	139
F.1. Ficha Técnica del MINTUR para Encuesta de Turismo Receptivo 2008	139
F.2. Matriz de Correlación para el análisis de datos sin agregar	141
F.3. Matriz de Saturación para el análisis de datos sin agregar	142
F.4. Matriz de Correlación para el análisis de datos agregados	142
F.5. Matriz de Saturación para el análisis de datos agregados	143
F.6. Historia de aglomeración según el algoritmo de Ward para el análisis de datos agregados	143
G: Apéndice de Códigos R	144
G.1. Series de Tiempo	144
G.2. Tipología de Turistas	149

ÍNDICE DE CUADROS

4.1. Modelo seleccionado ARIMA(1,0,0)(0,1,0)(4)	28
4.2. Predicción dentro de la muestra a 4 pasos para el modelo seleccionado ARIMA(1,0,0)(0,1,0)(4)	30
4.3. Predicción fuera de la muestra para el modelo Seleccionado ARIMA(1,0,0)(0,1,0)(4)	31
4.4. Comparación con igual trimestre del año anterior	32
4.5.a. Comparación de AIC entre modelos ARIMA(1,0,0)(0,1,1)(12)	39
4.5.b. Comparación de AIC entre modelos ARIMA(1,0,0)(1,1,0)(12)	39
4.6. Modelo seleccionado ARIMA(1,0,0)(0,1,1)(12)	39
4.7. Predicción dentro de la muestra a 12 pasos para el modelo seleccionado ARIMA(1,0,0)(0,1,1)(12)	43
4.8. Predicción dentro de la muestra para diciembre 2008, enero y febrero 2009	43
4.9. Predicción fuera de la muestra para el modelo seleccionado ARIMA(1,0,0)(0,1,1)(12)	45
4.10. Comparación con igual mes del año anterior	46
4.11. Comparación de RECM y AIC entre los ocho modelos	54
4.12. Modelo seleccionado ARIMA(0,0,1)(0,1,1)(4)	54
4.13. Predicción dentro de la muestra a 4 pasos para el modelo seleccionado ARIMA(0,0,1)(0,1,1)(4)	57
4.14. Predicción fuera de la muestra para el modelo seleccionado ARIMA(0,0,1)(0,1,1)(4)	58
4.15. Comparación con igual trimestre del año anterior	59
4.16. Número de grupos relevados por trimestre	65
4.17. Resumen de las variables	65
4.18. Valores de inercia acumulada	67
4.19. Salida del índice Calinski	69
4.20. Tamaño y composición de los clusters	69
4.21. Tabla de contingencia por modalidad de la variable Visito1 según los 5 clusters	72
4.22. Porcentajes fila de la tabla de contingencia por modalidad de la variable Visito1 según los 5 clusters	72
4.23. Porcentajes columna de la tabla de contingencia por modalidad de la variable Visito1 según los 5 clusters	73
4.24. Puntos de salida del grupo de viaje encuestado	75
4.25. Destino principal elegido por el grupo de viaje encuestado	75
4.26. Enumeración del punto de salida por trimestre	76
4.27. Reclasificación de la variable destino	77
4.28. Tabla de 14 variables utilizadas en el ACP	78
4.29. Tabla de inercia acumulada	78

4.30. Tamaño y composición de los clusters	84
4.31. Tabla de los puntos de salida por trimestre, según tipología de Método de Ward	88
4.32. Porcentajes fila de los puntos de salida por trimestre, según tipología de Método de Ward	89
4.33. Porcentajes columna de los puntos de salida por trimestre, según tipología de Método de Ward	89
4.34. Total expandido de turistas correspondiente a los puntos de salida por trimestre, según tipología de Método de Ward	90
4.35. Porcentajes fila del total expandido de turistas correspondiente a los puntos de salida por trimestre, según tipología de Método de Ward	90
4.36. Porcentajes columna del total expandido de turistas correspondiente a los puntos de salida por trimestre, según tipología de Método de Ward	90
4.37. Gasto total expandido correspondiente a los puntos de salida por trimestre, según tipología de Método de Ward	91
4.38. Porcentajes fila del gasto total expandido correspondiente a los puntos de salida por trimestre, según tipología de Método de Ward	91
4.39. Porcentajes columna del gasto total expandido correspondiente a los puntos de salida por trimestre, según tipología de Método de Ward	91
4.40. Comparación de los valores reales con las predicciones del modelo 1	95
4.41. Tasa de Crecimiento Real	95
C.1. Medidas de los residuos	129
C.2. Test de Ljung-Box	129
C.3. Test de Shapiro-Wilk y Jarque-Bera	129
C.4. Modelo descartado	129
C.5. Comparación de modelos	129
C.6. Predicción dentro de la muestra actualizando a un paso	130
C.7. Predicción dentro de la muestra actualizando a un paso	130
C.8. Predicción dentro de la muestra actualizando a un paso	130
C.9. Predicción dentro de la muestra actualizando a un paso	130
D.1. Medidas de los residuos	132
D.2. Test de Ljung-Box	132
D.3. Test de Shapiro-Wilk	132
D.4. Modelo descartado	132
D.5. Comparación de modelos	132
D.6. Predicción dentro de la muestra actualizando a un paso	133
D.7. Predicción dentro de la muestra actualizando a un paso	133
D.8. Predicción dentro de la muestra actualizando a un paso	133
D.9. Predicción dentro de la muestra actualizando a un paso	134
D.10. Predicción dentro de la muestra actualizando a un paso	134

D.11. Predicción dentro de la muestra actualizando a un paso	134
D.12. Predicción dentro de la muestra actualizando a un paso	134
D.13. Predicción dentro de la muestra actualizando a un paso	134
D.14. Predicción dentro de la muestra actualizando a un paso	134
D.15. Predicción dentro de la muestra actualizando a un paso	135
D.16. Predicción dentro de la muestra actualizando a un paso	135
E.1. Medidas de los residuos	137
E.2. Test de Ljung-Box	137
E.3. Test de Shapiro-Wilk y Jarque-Bera	137
E.4. Modelo descartado	137
E.5. Predicción dentro de la muestra actualizando a un paso	138
E.6. Predicción dentro de la muestra actualizando a un paso	138
E.7. Predicción dentro de la muestra actualizando a un paso	138
E.8. Predicción dentro de la muestra actualizando a un paso	138
F.1. Matriz de Correlación	141
F.2. Matriz de Saturación	142
F.3. Matriz de Correlación	142
F.4. Matriz de Saturación	143
F.5. Historia de aglomeración	143

ÍNDICE DE FIGURAS

4.1. Número de turistas ingresados a Uruguay por trimestre en el período 1993.I-2009.I	22
4.2. Número de turistas ingresados por año en el período 1993.I-2009.I	23
4.3. Número de turistas ingresados por trimestre en el período 1993.I-2009.I	23
4.4. Autocorrelograma simple y parcial del Modelo 1	24
4.5. Gráfico y autocorrelograma simple y parcial de la primera diferencia estacional del logaritmo del Modelo 1 - $\Delta_4 \log(\text{seriet})$	25
4.6. Predicción dentro de la muestra a 4 pasos para el modelo seleccionado ARIMA(1,0,0)(0,1,0)(4)	31
4.7. Predicción fuera de la muestra para el modelo seleccionado ARIMA(1,0,0)(0,1,0)(4)	32
4.8. Número de turistas ingresados a Uruguay por mes en el período 1996.01-2009.07	33
4.9. Número de turistas ingresados por año en el período 1996.01-2009.07	34
4.10. Número de turistas ingresados por mes en el período 1996.01-2009.07	34
4.11. Autocorrelograma simple y parcial del Modelo 2	35
4.12. Gráfico y autocorrelograma simple y parcial de la primera diferencia estacional del logaritmo del Modelo 2 - $\Delta_{12} \log(\text{seriem})$	36
4.13. Predicción dentro de la muestra a 12 pasos para el modelo seleccionado ARIMA(1,0,0)(0,1,1)(12)	44
4.14. Predicción fuera de la muestra para el modelo seleccionado ARIMA(1,0,0)(0,1,1)(12)	45
4.15. Número de turistas uruguayos residentes en el exterior ingresados a Uruguay por trimestre en el período 1996.I-2009.I	47
4.16. Número de turistas uruguayos residentes en el exterior ingresados por año en el período 1996.I-2009.I	48
4.17. Número de turistas uruguayos residentes en el exterior ingresados por trimestre en el período 1996.I-2009.I	48
4.18. Autocorrelograma simple y parcial del Modelo 3	50
4.19. Gráfico y autocorrelograma simple y parcial de la primera diferencia estacional del logaritmo del Modelo 3 - $\Delta_4 \log(\text{seried})$	51
4.20. Predicción dentro de la muestra a 4 pasos para el modelo seleccionado ARIMA(0,0,1)(0,1,1)(4)	57
4.21. Predicción fuera de la muestra para el modelo seleccionado ARIMA(0,0,1)(0,1,1)(4)	59
4.22. Mapa de Uruguay indicando los puntos de salida	61
4.23. Calidad de representación de las variables en el plano factorial. (F_1 y F_2)	68

4.24. Diagrama de cajas por cluster para los tres componentes principales	70
4.25. Calidad de representación de las variables en el plano principal	81
4.26. Proyección de los individuos en el plano principal	82
4.27. Dendrograma según algoritmo de Ward	84
4.28. Diagrama de cajas por cluster para los cinco componentes principales, según el algoritmo de Ward	85
4.29. Histograma de los puntos de salida por trimestre	89
C.1. Gráfico y autocorrelograma simple y parcial de los residuos para el modelo 1	128
D.1. Gráfico y autocorrelograma simple y parcial de los residuos para el modelo 2	131
E.1. Gráfico y autocorrelograma simple y parcial de los residuos para el modelo 3	136

CAPÍTULO 1: INTRODUCCIÓN

El turismo es uno de los factores que contribuye a la generación del producto de una economía nacional. Muchos gobiernos han reconocido el rol primordial que juega el turismo en el crecimiento económico y han desarrollado su potencial turístico rápidamente. Se puede señalar que mientras en 1950 el turismo internacional recibía U\$S 2,1 mil millones, en 2004 había crecido a U\$S 622,7 mil millones, según Brida *et al.* (2008) [1].

En el presente trabajo se buscaron dos objetivos. Por un lado se desarrollaron modelos que permiten predecir el número de turistas que ingresarán a Uruguay con frecuencia trimestral y mensual. Por otro lado se realizó análisis de las tipologías de los turistas en base a la Encuesta de Turismo Receptivo 2008 recabada por el Ministerio de Turismo y Deporte del Uruguay.

Para el primer objetivo se utilizaron técnicas estadísticas para generar modelos de análisis y predicción que puedan convertirse en herramientas de planificación y toma de decisiones tanto para agentes del sector público como del privado.

Para el segundo objetivo se consideró importante realizar un análisis de los turistas que visitaron Uruguay en el año 2008. En este sentido se planteó la necesidad de agregar al análisis de la cantidad de turistas una descripción de sus principales características como por ejemplo: destino, tiempo medio de estadía en el país y composición del gasto de los turistas que visitaron Uruguay durante el año 2008.

Se contó con información para un solo año (2008), lo que es una limitante, ya que si el trabajo se hubiera realizado con información de otro año las conclusiones hubieran sido otras. Como consecuencia, se considera que el tratamiento de la información es lo que resulta clave para el desarrollo de políticas públicas y privadas en materia de turismo.

La profundización del conocimiento del turista que ingresa a Uruguay contribuye a desarrollar estrategias de segmentación adecuadas que permiten generar productos turísticos atractivos.

Se entiende que este trabajo puede ser de interés para el Ministerio de Turismo y los distintos agentes privados de la industria turística en tanto les brindará herramientas

para el tratamiento de nueva información relevante para la construcción de sus estrategias.

En el capítulo 2 se exhibieron los objetivos del proyecto y en el 3 los antecedentes, es decir trabajos anteriores en materias referidas a nuestro tema de estudio. En el capítulo 4 se presentaron los resultados obtenidos y en el 5 las conclusiones y recomendaciones.

CAPÍTULO 2: OBJETIVOS

El trabajo se basó en la estimación del número de turistas que ingresarán a Uruguay y en la caracterización de su perfil para los ingresados en el año 2008.

Para cumplir con el primer objetivo se aplicaron los siguientes modelos SARIMA:

1. Modelo univariado de series de tiempo que se ajuste al número de turistas que ingresaron a Uruguay trimestralmente, en el período 1993.I a 2009.I.
(Modelo 1)
2. Modelo univariado de series de tiempo que se ajuste al número de turistas que ingresaron a Uruguay mensualmente, en el período 1996.01 a 2009.07.
(Modelo 2)
3. Modelo univariado de series de tiempo que se ajuste al número de turistas uruguayos residentes en el exterior que ingresaron a Uruguay trimestralmente, en el período 1996.I a 2009.I. (Modelo 3)

Cabe señalar que la serie trimestral para el período 1993.I a 2009.I del Modelo 1 incluye a los extranjeros que perciben algún tipo de remuneración en Uruguay y a los turistas uruguayos residentes en el exterior que ingresan a Uruguay, mientras que la serie mensual 1996.01 a 2009.07 del Modelo 2 no los incluye. Se generó una tercera serie trimestral que abarca el período 1996.I – 2009.I, Modelo 3, que surgió de la diferencia algebraica entre los modelos 1 y 2. Se aclaró que como los extranjeros que perciben algún tipo de remuneración en Uruguay son muy pocos, se supuso que la diferencia se debió mayoritariamente a los turistas uruguayos residentes en el exterior.

Para los tres modelos se obtuvieron predicciones a un año dentro y fuera de la muestra.

Las fuentes de información para la primera etapa fueron:

- La serie trimestral 1993.I-2009.I fue proporcionada por el Ministerio de Turismo y Deporte del Uruguay.
- La serie mensual 1996.01- 2009.07 fue proporcionada por la Dirección Nacional de Migraciones a través del Ministerio de Turismo y Deporte del Uruguay.

En los tres modelos se incluyeron los visitantes que permanecieron menos de 24 horas o estuvieron en tránsito a otro país. No se incluyeron los visitantes de cruceros ni quienes no hacen trámite migratorio (en el caso de la frontera seca de nuestro territorio, por ejemplo Rivera - Livramento, Chuy - Chui, Río Branco - Yaguarón)

En el segundo objetivo, para la caracterización del perfil del turista se utilizaron técnicas de análisis multivariado (Análisis de Clusters y Análisis de Componentes Principales) con el fin de obtener tipologías de los turistas ingresados a Uruguay en el año 2008. Se plantearon dos análisis, uno con el objetivo de encontrar tipologías de los turistas que ingresaron a Uruguay durante el año 2008 y otro con el objetivo de analizar la evolución temporal de los puntos de salida en los que son encuestados los grupos de viaje a lo largo de los cuatro trimestres del 2008.

La fuente de información para la segunda etapa surgió de la Encuesta de Turismo Receptivo del año 2008 realizada por el Ministerio de Turismo y Deporte del Uruguay.

En estos datos se incluyeron los turistas uruguayos residentes en el exterior y los visitantes que permanecieron menos de 24 horas o estuvieron en tránsito a otro país. No se incluyeron los visitantes de cruceros ni quienes no hacen trámite migratorio (en el caso de la frontera seca de nuestro territorio, por ejemplo Rivera - Livramento, Chuy - Chui, Río Branco - Yaguarón).

CAPÍTULO 3: ANTECEDENTES

Se encontraron las siguientes investigaciones vinculadas al análisis y modelización de la evolución temporal del turismo, las cuales se detallan a continuación.

En la primera de ellas, realizada en España por Espasa *et al.* (1993) [2], se encontró un análisis econométrico que estudió el turismo receptivo en España y a partir de los resultados se obtuvieron conclusiones sobre los determinantes del sector y las medidas que se podrían tomar para reactivarlo. Utilizaron como variable explicativa un índice que refleja la renta real agregada de los países clientes y dos índices de precios relativos; uno que toma en cuenta los países competidores y otro para los países clientes o demandantes de turismo. La periodicidad de los datos es trimestral, siendo el período considerado 1978-1990.

En una investigación realizada en Uruguay por Robano (2000) [3] se analizaron los determinantes del gasto realizado por los turistas que ingresaron al país, utilizando datos trimestrales entre 1987 y 2000, empleando técnicas de cointegración. Sus resultados indicaron la existencia de una relación estable a largo plazo entre las exportaciones de servicios turísticos y el ingreso (o consumo) argentino y los precios relativos de Argentina con Uruguay.

En otro orden, Armellini e Isabella (2003) [4] buscaron determinar la contribución del turismo receptivo al valor agregado, al empleo y a las remuneraciones, en el período comprendido entre 1996 y 2002. La metodología que utilizó este trabajo descansó básicamente en las técnicas de la contabilidad nacional. El trabajo concluyó que el turismo receptivo es la principal fuente de demanda externa en términos de generación de valor agregado y remuneraciones y una de las principales fuentes generadoras de empleo.

Otro trabajo es el Memorando realizado por Sosa y Glejberman (2003) [5], en el mismo se presentaron cuatro proyecciones, una de ellas sin tomar en cuenta variables predictoras, y las tres restantes correspondientes a tres escenarios posibles para la variable predictora. La variable predictora que utilizaron es el indicador de competitividad y los tres escenarios diseñados fueron denominados: pesimista, normal y optimista, los cuales supusieron pérdida de competitividad de la economía uruguaya respecto de la de sus vecinos, competitividad constante y ganancia de competitividad, respectivamente. El indicador de competitividad tomó en cuenta la relación entre el

tipo de cambio y los precios internos de Argentina, Brasil y Uruguay. Este indicador consiste en deflactar el IPC de cada país entre un índice del Tipo de Cambio (ITC). Se otorga mayor ponderación a Argentina porque es el país que tiene mayor peso en la demanda del turismo receptivo.

Otro estudio realizado es el de Mantero *et al.* (2004) [6], cuyo objetivo fue estimar la demanda de turismo receptivo en Uruguay, a partir de datos mensuales entre 1994.01 y 2004.01, utilizando técnicas de cointegración. Dado que el comportamiento de los turistas fue heterogéneo, se estimaron dos tipos de modelos, uno con datos agregados y otro que tomó en consideración la nacionalidad de los turistas (argentinos, brasileños, chilenos y otras nacionalidades). Ellos encontraron que la desagregación por nacionalidad aportó información relevante para entender la evolución pasada del turismo global y constituyó una mejor aproximación estadística a la estimación del número de turistas totales. Los determinantes del ingreso de turistas variaron de acuerdo a la nacionalidad (nivel de actividad de los respectivos países, precios relativos de la economía uruguaya respecto de los precios relativos del país de origen, precios relativos respecto de destinos competitivos), reflejando el comportamiento heterogéneo en el agregado de turistas.

Brida *et al.* (2008) [1] analizaron los efectos de la actividad turística sobre el crecimiento económico uruguayo, en particular la actividad turística asociada a los visitantes argentinos ya que Argentina es el principal emisor de turismo a Uruguay, utilizando para ello información trimestral para el período 1987-2006. El análisis de cointegración mostró la existencia de un vector de cointegración entre el PIB per cápita, el gasto de los turistas argentinos y el tipo de cambio bilateral entre Uruguay y Argentina. Por su parte, se sugirió que la dirección de la causalidad es desde el gasto real de los turistas al PIB per cápita.

Otro artículo consultado fue realizado por Gil-Alana *et al.* (2008) [7]. El mismo analizó el número de turistas que ingresaron mensualmente a las Islas Canarias (1992.01-2005.12) utilizando distintos modelos de series de tiempo estacionales univariados. Se consideró estacionalidad determinística y estocástica. Para esta última se emplearon modelos estacionales de raíz unitaria y modelos integrados estacionales. También se empleó un modelo con distintos órdenes de integración en las frecuencias estacionales. Estos modelos fueron comparados en términos de su habilidad predictiva fuera de la muestra. Los resultados mostraron que un modelo determinístico simple AR(1) con variables dummies estacionales tuvo mejores resultados en el corto plazo

que otros modelos basados en diferenciación estacional fraccional o entera. Sin embargo, en el largo plazo los resultados no distinguieron entre el modelo basado en variables dummies estacionales y otro usando integración fraccional estacional.

Para la caracterización del perfil del turista que ingresa a Uruguay se encontró un trabajo realizado por Glejberman, (2004) [8]. El objeto de este informe fue identificar tipologías de visitantes y sus principales características, principalmente en relación con el gasto que realizan, a los efectos de proporcionar información relevante para la definición de las estrategias de promoción.

Las tipologías se definieron a priori, a partir de la intersección de ciertas variables de la Encuesta de Turismo Receptivo con datos del primer trimestre de 2004. Mediante tales variables se crearon escenarios en base a información tal que se cuantificó su importancia en relación con el total de visitantes. Las variables utilizadas para este propósito fueron: motivo del viaje, sexo y edad de los visitantes, número de integrantes del grupo y tiempo de estadía.

Cabe mencionar el artículo de Brida *et al.* (2010) [9], cuyo objetivo fue estudiar el perfil de turistas relacionado a los distintos gastos hechos por los visitantes que llegaron en cruceros a Cartagena de Indias. Se consideraron cuatro categorías de gasto: gasto turístico, comida y bebida, souvenirs y joyas. En el análisis de regresión utilizaron dos modelos para explicar el comportamiento del gasto de los cruceristas. En primer lugar, se utilizó un modelo para determinar las variables que explican de mejor manera la decisión de gasto. En segundo lugar, se utilizó un modelo con variable dependiente censurada para intentar determinar los efectos de distintas variables sobre el nivel de gasto de cada categoría. Se utilizó el modelo Tobit, ya que esta metodología permitió utilizar toda la información disponible sobre la variable dependiente al incorporar de manera conjunta la decisión de gastar o no gastar y cuanto gastar. Por lo tanto, la decisión de gasto en las cuatro categorías es considerada como una decisión binaria. A partir de esta especificación de la variable dependiente, se utilizó el modelo Logit para determinar las variables que mejor explican la decisión de gasto de cada categoría. Las técnicas basadas en la segmentación del gasto fueron usadas para explorar si los cruceristas que difieren significativamente en su gasto pudieron ser identificados por algunas características.

Los hallazgos de este estudio mostraron que el segmento de grandes consumidores puede ser distinguido del resto de los cruceristas por las siguientes características: a) gastan más en joyas y en general gastan más en todas las categorías; b) son más jóvenes; c) tienen ingresos altos; d) los grandes consumidores en joyas son mujeres

residentes en Estados Unidos; e) están viajando por primera o segunda vez en un crucero; f) permanecieron un período más extenso en Cartagena de Indias; g) tienden a viajar en grupos.

CAPÍTULO 4: RESULTADOS

Los gráficos, las estimaciones y los contrastes estadísticos que se presentan fueron realizados con el programa R 2.9.2 y con Excel.

4.1. Modelización de la serie trimestral 1993.I – 2009.I - Modelo 1

4.1.1. Datos

La serie que se presenta corresponde al ingreso de turistas a Uruguay en el período Enero 1993 a Marzo 2009. Los datos tienen periodicidad trimestral y fueron proporcionados por el Ministerio de Turismo y Deporte del Uruguay.

4.1.2. Análisis descriptivo de la serie

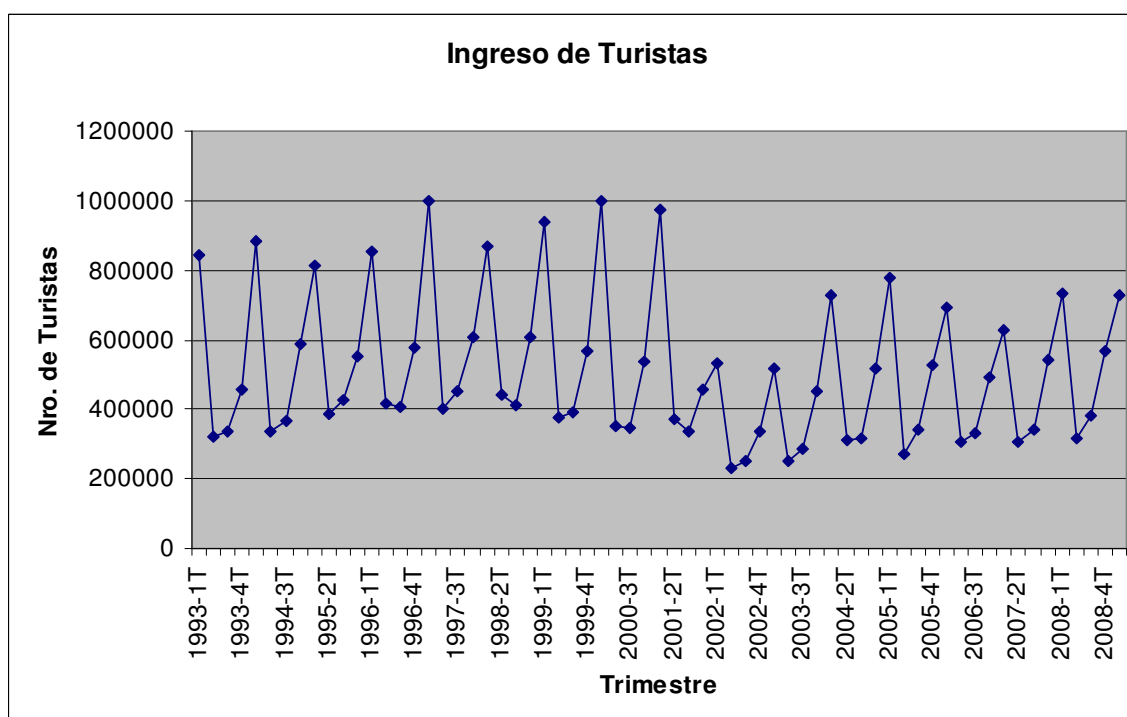


Figura 4.1. Número de turistas ingresados a Uruguay por trimestre en el período 1993.I-2009.I.

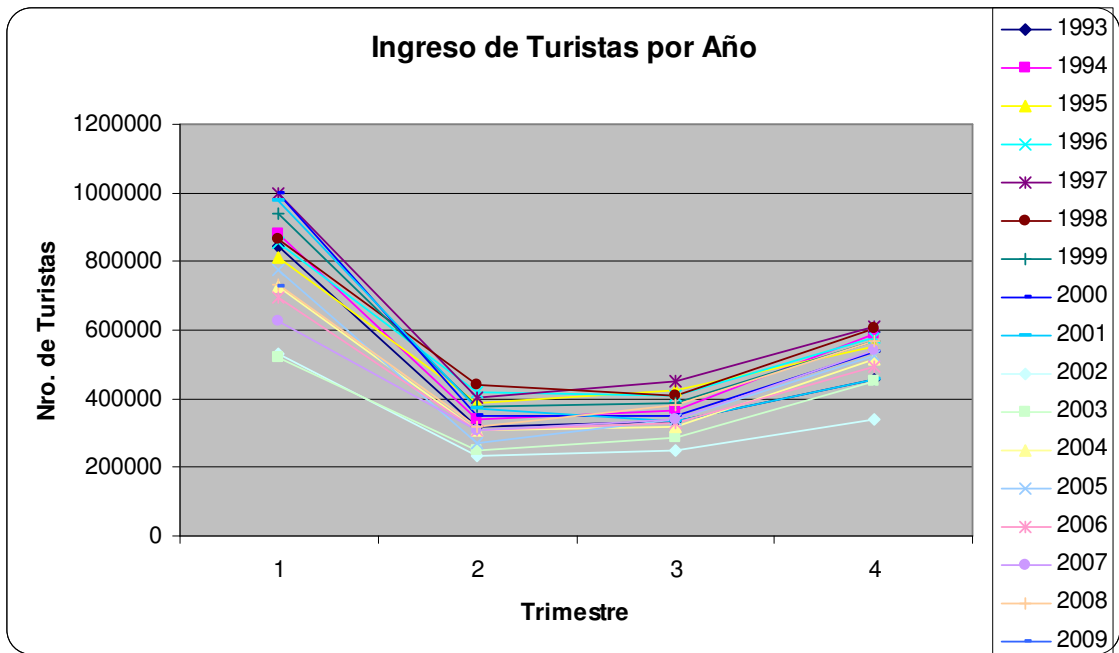


Figura 4.2. Número de turistas ingresados por año en el período 1993.I-2009.I.

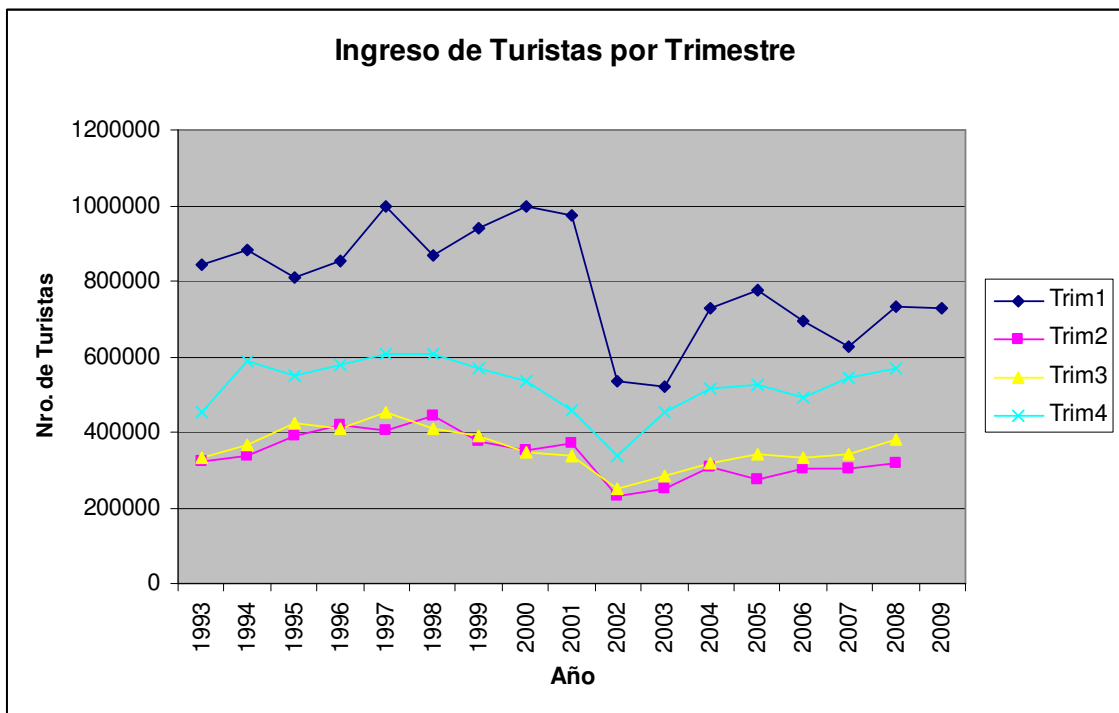


Figura 4.3. Número de turistas ingresados por trimestre en el período 1993.I-2009.I.

La figura 4.1 corresponde al número de turistas ingresados a Uruguay por trimestre. En ella se observó una fuerte estacionalidad en los meses de verano. Se observó una fuerte caída en el ingreso de turistas a partir del primer trimestre del 2002, sin volver a alcanzar nunca los niveles de los trimestres anteriores.

La figura 4.1 reflejó la existencia de una serie con media y varianza no constantes lo que indicó que el proceso es no estacionario. Por tanto, la serie debió ser transformada para ser estacionaria. Para esto se aplicó logaritmo a la serie y luego se aplicó una diferencia estacional (Δ_4).

La figura 4.2 correspondiente al número de turistas ingresados a Uruguay por año también sugirió una fuerte estacionalidad ya que en todos los años se observó un comportamiento similar a lo largo de los trimestres; en el primer trimestre ingresó el mayor número de turistas, seguido por el cuarto trimestre y luego alternaron entre el segundo y tercer trimestre.

En la figura 4.3 se comprobó lo explicado anteriormente al observar que el número de turistas ingresados al Uruguay fue mayor para los primeros trimestres de cada año, seguidos por el cuarto trimestre y finalmente segundo y tercer trimestre.

En base a las figuras 4.1, 4.2 y 4.3 se pudo ver que la serie no tiene tendencia.

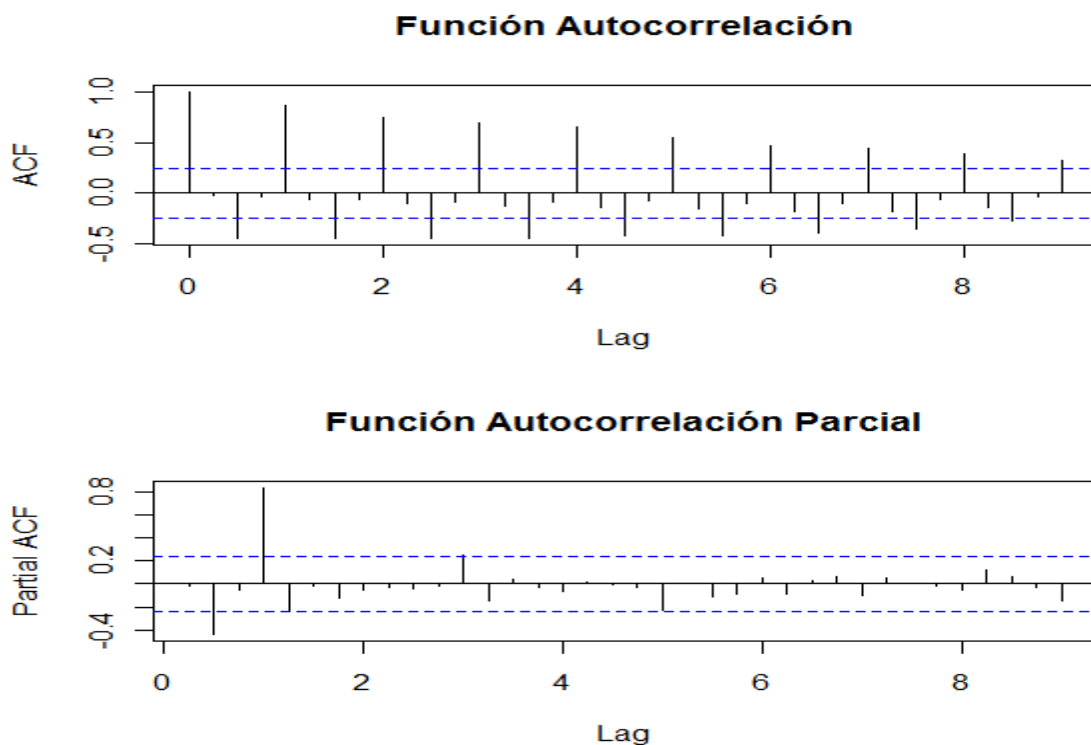


Figura 4.4. Autocorrelograma simple y parcial del Modelo 1.

En la figura 4.4 se comprobó la no estacionariedad de la serie ya que el autocorrelograma simple no decayó rápidamente y además mostró cambio de signo.

En el autocorrelograma parcial se observó que dos valores significativos sobresalieron de la banda de confianza, los retardos 2 y 4. Se observó con claridad la estacionalidad de la serie en la función de autocorrelación.

En base a lo observado, para lograr estacionariedad de la serie, se decidió aplicar a la serie una diferencia estacional, para extraer la estacionalidad de la serie, además de la transformación logarítmica previamente aplicada.

A continuación se presenta como quedó la serie transformada.

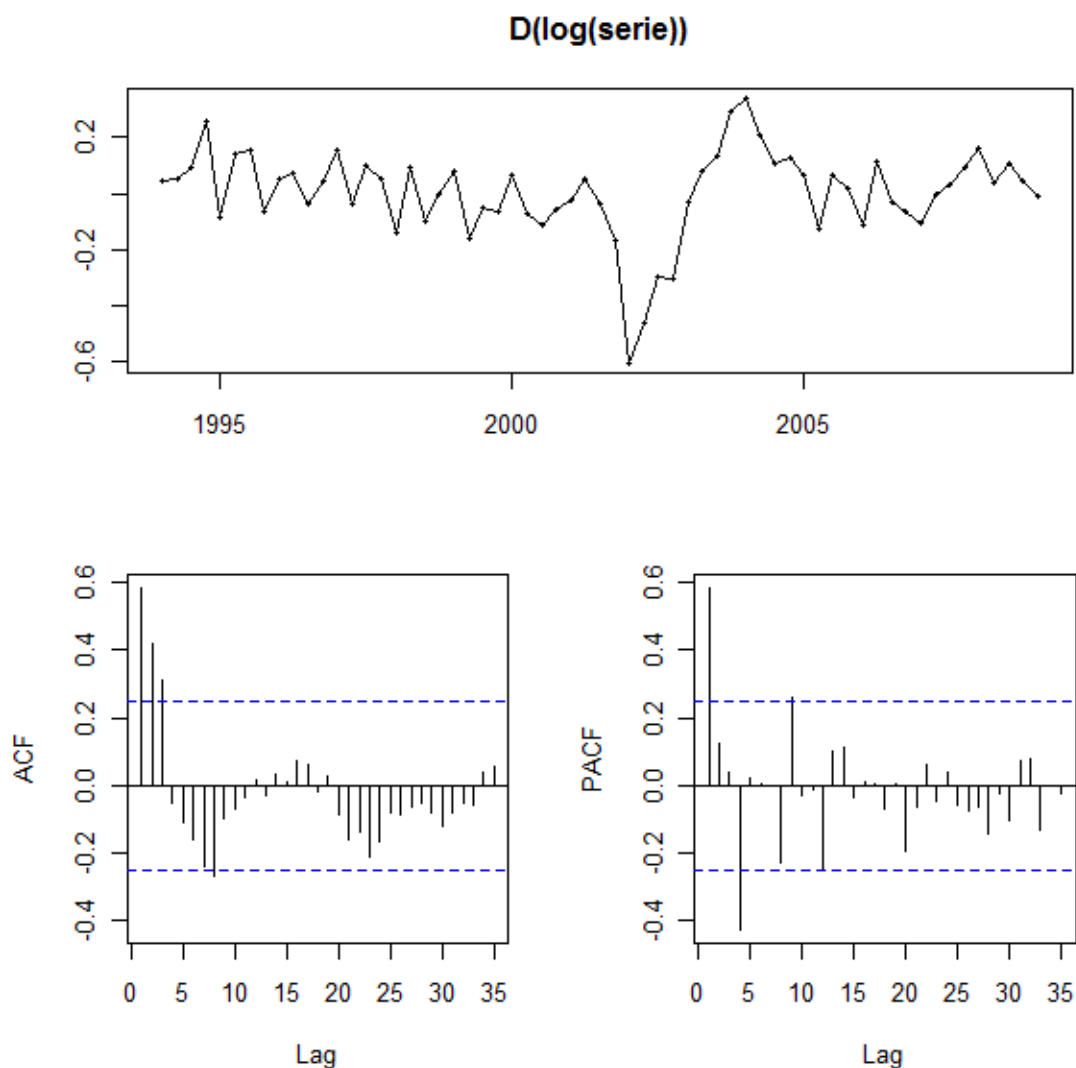


Figura 4.5. Gráfico y autocorrelograma simple y parcial de la primera diferencia estacional del logaritmo del Modelo 1 - $\Delta_4 \log(\text{serie})$.

Observando la figura 4.5 se concluyó que la nueva serie transformada es estacionaria en media y varianza, ya que osciló en torno a su nivel medio. Hubo una fuerte caída en

el año 2002 que se debió a la incidencia de la crisis económica regional de fines del 2001 que repercutió fuertemente en el número de turistas ingresados al país a partir del primer trimestre del año 2002 y los siguientes trimestres.

4.1.3. Ajuste del modelo

Luego de haber analizado los gráficos de la serie original y de la serie transformada (con logaritmo y una diferencia estacional) se probó con los siguientes modelos para lograr el ajuste de los datos.

$$\begin{array}{ll}
 \text{ARIMA } (1,0,0) (0,1,1)_{(4)} & \text{ARIMA } (0,0,1) (1,1,0)_{(4)} \\
 \text{ARIMA } (1,0,0) (0,1,0)_{(4)} & \text{ARIMA } (0,0,1) (0,1,1)_{(4)} \\
 \text{ARIMA } (0,0,1) (0,1,0)_{(4)} & \text{ARIMA } (1,0,0) (1,1,0)_{(4)}
 \end{array}$$

A su vez, a cada modelo se le aplicaron las siguientes intervenciones:

- **Turismo:** Se consideró la variable indicatriz Turismo para indicar en que trimestre del año se ubicó la semana de turismo. Puede ser en el primer o segundo trimestre de cada año o en ambos.
- **Tc2002:** Se consideró como cambio transitorio la crisis del 2002 y su efecto en el número de turistas que ingresaron al país.
- **Tc2006:** Se consideró como cambio transitorio al impacto debido al corte de los puentes entre Argentina y Uruguay que comenzó en Diciembre 2005 y afectó el número de turistas que ingresaron al país provenientes de Argentina a partir del primer trimestre del 2006.
- **Ls2002:** Se consideró como cambio de nivel la crisis del 2002 y su efecto en el número de turistas que ingresaron al país.
- **Ls2006:** Se consideró como cambio de nivel al impacto debido al corte de los puentes entre Argentina y Uruguay que comenzó en Diciembre 2005 y afectó el número de turistas que ingresaron al país provenientes de Argentina a partir del primer trimestre del 2006.

Se consideraron las intervenciones mencionadas alternativamente de la siguiente manera:

- Turismo , Tc2002, Tc2006
- Turismo, Tc2002
- Turismo, Ls2002, Ls2006
- Turismo, Ls2002

De acuerdo a los resultados obtenidos, los siguientes modelos intermedios para el ajuste del número de turistas que ingresan a Uruguay cumplieron con los requisitos necesarios para la significación de los coeficientes estimados:

- ARIMA (1,0,0) (0,1,0)₍₄₎ con Turismo y Tc2002.
- ARIMA (1,0,0) (0,1,0)₍₄₎ con Turismo y Ls2002.

Para predecir el número de turistas que ingresan al país resultaron significativos el efecto de la semana de turismo y el efecto de la crisis del 2002, que puede considerarse como un cambio transitorio o como un cambio de nivel, es decir, su efecto va decreciendo a medida que pasa el tiempo o su efecto se mantiene hasta el final de la serie analizada.

No resultaron significativos para nuestro análisis el efecto del corte de los puentes del 2006, ya sea considerado como cambio transitorio o de nivel, es decir Tc2006 y Ls2006.

4.1.4. Modelo seleccionado

Se estimaron los coeficientes para estos dos modelos y en ambos casos se obtuvieron coeficientes significativos.

Sin embargo, se descartó el modelo ARIMA (1,0,0) (0,1,0)₍₄₎ con Turismo y Ls2002¹¹ ya que a la hora de realizar predicciones este modelo predijo con mayor error (mayor RECM y mayor EAM)¹².

El modelo seleccionado es ARIMA (1,0,0) (0,1,0)₍₄₎ con Turismo y Tc2002.

¹¹ El modelo descartado está desarrollado en el Apéndice C.2.

¹² Ver Apéndice C.3.

ARIMA (1,0,0) (0,1,0) ₍₄₎	Coefficientes	t observados
Ar1	0,53	4,93
Turismo	0,08	4,97
Tc2002	-0,43	-6,41

Cuadro 4.1. Modelo seleccionado ARIMA(1,0,0)(0,1,0)(4).

Se observó que los coeficientes del modelo estimado son significativos al 5%.

Se verificó que este modelo cumplió con los supuestos de los residuos.

El cumplimiento de los supuestos de los residuos verifica los siguientes aspectos:

- Media cero:

En la figura C.1 del apéndice C se observó que los residuos fluctúan en torno a la media que se aproxima a cero. También se observó que las barras de los coeficientes de autocorrelación de los residuos permanecieron dentro de las bandas de confianza. Ver cuadro C.1 de las medidas de resumen de los residuos en el apéndice C.

- Incorrección:

Para testear la incorrección de los residuos se utilizó el test de Ljung-Box. Dicha prueba se realizó tomando $k = 4, 8, 12$ y 16 y resultó positiva para aceptar la hipótesis nula de ausencia de autocorrelación en los residuos, al 5%.

Se corroboró que los residuos están incorreccionados al observar que las barras de los coeficientes de autocorrelación de los residuos permanecieron dentro de las bandas de confianza, evidenciando que la dinámica de la serie fue captada por la modelización propuesta. Ver cuadro C.2 del test de Ljung-Box en el apéndice C.

- Normalidad:

Para testear la normalidad de los residuos se utilizó el test de Shapiro-Wilk y el test de Jarque-Bera. Dichos tests llevaron a no rechazar la hipótesis nula de normalidad de los residuos, al 5%. Ver cuadro C.3 del test de Shapiro-Wilk y Jarque-Bera en el apéndice C.

4.1.5. Interpretación del modelo seleccionado

El modelo seleccionado se planteó de la siguiente manera:

$$\log(\hat{y}_t) = 0,53 \log(y_{t-1}) + \log(y_{t-4}) - 0,53 \log(y_{t-5}) + 0,08Turismo - 0,43Tc2002$$

Siendo:

- Turismo la indicatriz que toma valor 1 en los trimestres del año en que cae la semana de Turismo y 0 en el resto.

$$I_t^{Turismo} = \begin{cases} 1, t = Turismo \\ 0, t \neq Turismo \end{cases}$$

- Tc2002 el cambio transitorio de la crisis económica regional que comenzó en Diciembre 2001 y que afectó a la baja el ingreso de turistas a Uruguay a partir del primer trimestre del año 2002 en adelante. Como su efecto es transitorio, el mismo va disminuyendo a medida que transcurre el tiempo hasta desaparecer. Toma valor 1 en el primer trimestre 2002 y luego disminuye de la siguiente manera:

$$\frac{1}{1-0,7L} = (1 + 0,7 + 0,7^2 + 0,7^3 + \dots)$$

La estimación y predicción en logaritmos en el momento t depende de:

- La observación del trimestre anterior en logaritmos $\log(y_{t-1})$, cuyo peso en la generación de la nueva observación es de 0,53.
- La observación en logaritmos de cuatro trimestres atrás $\log(y_{t-4})$, con peso 1.
- La observación en logaritmos de cinco trimestres atrás $\log(y_{t-5})$, con peso - 0,53.

Es decir, $\log(\hat{y}_t)$ depende de la observación del trimestre anterior, la observación del mismo trimestre correspondiente al año anterior y el efecto del año anterior más un trimestre. Cada una de estas observaciones con sus respectivos pesos.

A esto se le agrega el efecto del análisis de intervención:

- Turismo es una variable indicatriz que toma valor uno en los trimestres en que cae la semana de turismo y cero en el resto de los trimestres. Tiene peso 0,08.
- Tc2002 es el cambio transitorio que considera a la crisis económica del 2002 y su efecto en el número de turistas que ingresan al país es -0,43. Como ya ha sido explicado, al ser un cambio transitorio el efecto de la crisis del 2002 va disminuyendo a medida que transcurre el tiempo.

Se interpretó que la predicción del número de turistas que ingresará a Uruguay depende del número de turistas que haya ingresado un trimestre y un año atrás con efecto positivo y se notó que incide más el número de turistas que ingresaron un año atrás que un trimestre atrás debido a que esta serie es claramente estacional. El número de turistas que ingresaron cinco trimestres antes tiene un efecto negativo en la predicción del número de turistas que ingresará a Uruguay.

También se interpretó que el trimestre en que cae la semana de turismo tiene un leve efecto positivo y la crisis del 2002 tiene un efecto negativo decreciente en la predicción del número de turistas que ingresará a Uruguay.

4.1.6. Predicción

4.1.6.1. Predicción dentro de la muestra

Para probar el modelo estimado se cortó la serie y se generaron predicciones dentro de la muestra para dicho intervalo.

Trimestre	Predicción	Valor Real	Int. Inf. 95	Int. Sup. 95	Error	Error en %
2T-2008	290.302	317.096	241.675	348.714	26.794	8,4%
3T-2008	348.137	380.560	283.166	428.015	32.423	8,5%
4T-2008	547.689	566.788	442.888	677.288	19.099	3,4%
1T-2009	674.461	726.441	544.562	835.347	51.980	7,2%

Cuadro 4.2. Predicción dentro de la muestra a 4 pasos para el modelo seleccionado ARIMA(1,0,0)(0,1,0)(4).

Para evaluar las predicciones se planteó un nivel de confianza del 95%. Para un nivel de confianza mayor se pierde precisión ya que los intervalos de confianza generados

son más amplios. Observando el cuadro 4.2 se concluyó que los valores reales de la serie estaban dentro del intervalo de confianza para los 4 períodos.

Para cada caso, se comparó el valor real observado con la predicción del modelo calculando el error y luego el error relativo porcentual de cada predicción.

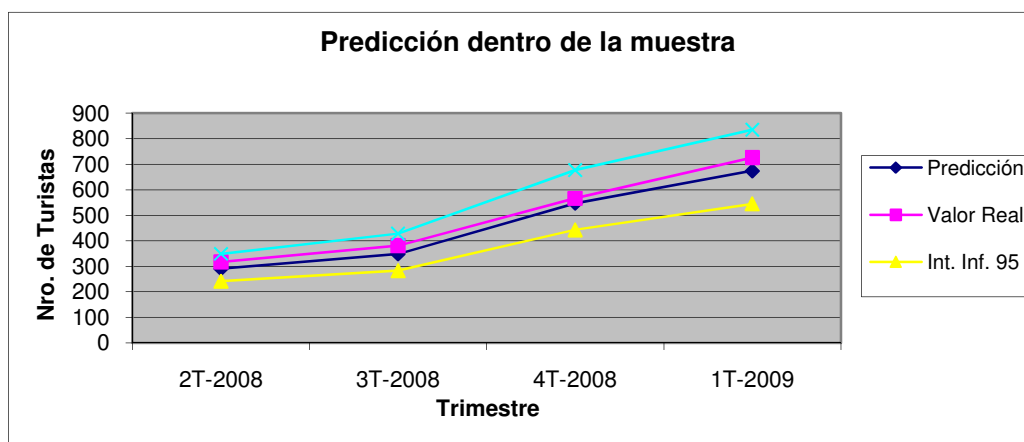


Figura 4.6. Predicción dentro de la muestra a 4 pasos para el modelo seleccionado ARIMA(1,0,0)(0,1,0)(4).

Posteriormente se realizaron las predicciones dentro de la muestra para el modelo seleccionado actualizando a un paso, es decir incluyendo un valor observado real con el objetivo de comprobar si las predicciones mejoraban al incorporar observaciones reales.

Observando los cuadros C.6, C.7, C.8 y C.9 del apéndice C se concluyó que los valores reales de la serie estaban dentro del intervalo de confianza al 95% para todos los períodos, lo cual indicó que las predicciones realizadas por el modelo seleccionado fueron correctas.

Para cada caso, se comparó el valor real observado con la predicción del modelo calculando el error y luego el error relativo porcentual de cada predicción.

Se observó que a medida que se fueron incorporando observaciones reales al modelo estimado, las predicciones mejoraban, ya que los errores fueron disminuyendo.

4.1.6.2. Predicción fuera de la muestra

Trimestre	Predicción	Int. Inf. 95	Int. Sup. 95
2T-2009	358.618	299.762	429.030
3T-2009	388.732	317.182	476.421
4T-2009	573.284	464.770	707.135
1T-2010	794.501	642.947	981.779

Cuadro 4.3. Predicción fuera de la muestra para el modelo seleccionado ARIMA(1,0,0)(0,1,0)(4).

Predicción fuera de la muestra

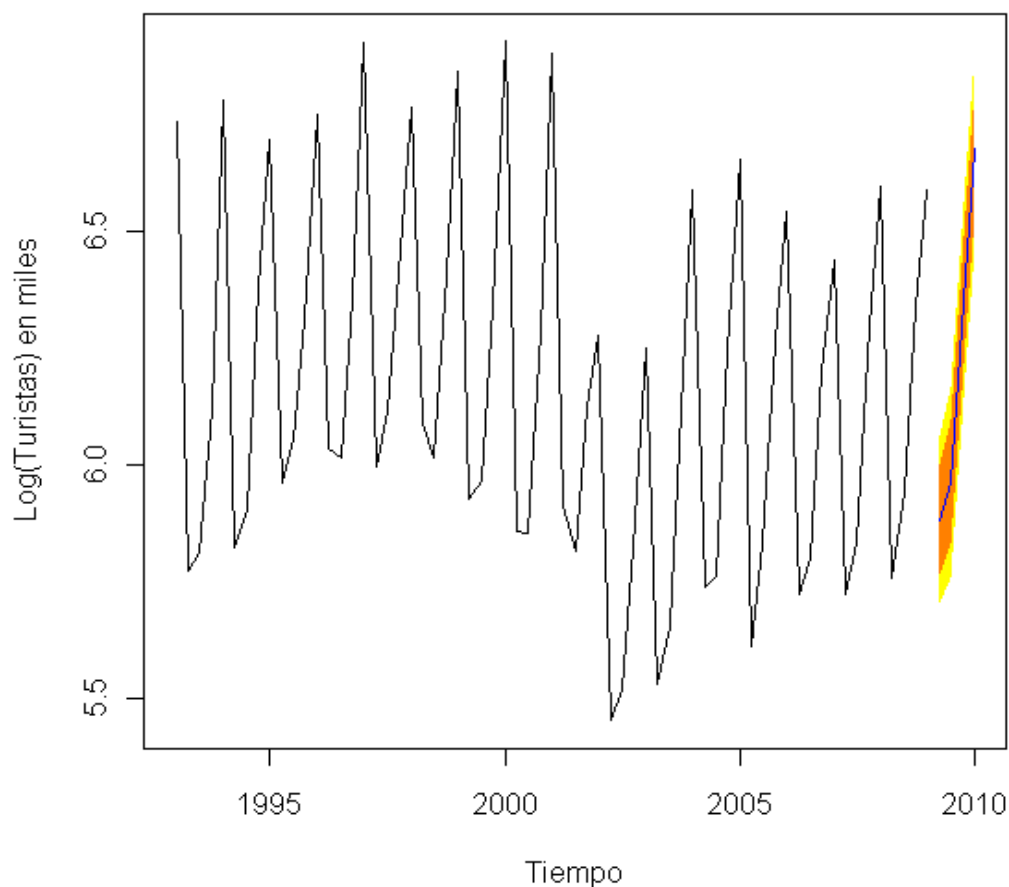


Figura 4.7. Predicción fuera de la muestra para el modelo seleccionado $ARIMA(1,0,0)(0,1,0)(4)$.

Trimestre	2008/2009	2009/2010	Tasa de crecimiento
2T	317.096	358.618	13,09%
3T	380.560	388.732	2,15%
4T	566.788	573.284	1,15%
1T	726.441	794.501	9,37%

Cuadro 4.4. Comparación con igual trimestre del año anterior.

Observando el cuadro 4.4 se pudo pronosticar que es de esperar que el número de turistas que ingresará a Uruguay en los próximos cuatro trimestres sea superior a la cantidad de turistas ingresados en los últimos cuatro trimestres (2T08, 3T08, 4T08 y 1T09).

4.2. Modelización de la serie mensual 1996.01 – 2009.07 – Modelo 2

4.2.1. Datos

La serie que se presenta corresponde al ingreso de turistas a Uruguay en el período Enero 1996 a Julio 2009. Los datos tienen periodicidad mensual y fueron proporcionados por el Ministerio de Turismo y Deporte del Uruguay.

En esta serie no están incluidos los turistas uruguayos residentes en el exterior.

4.2.2. Análisis descriptivo de la serie

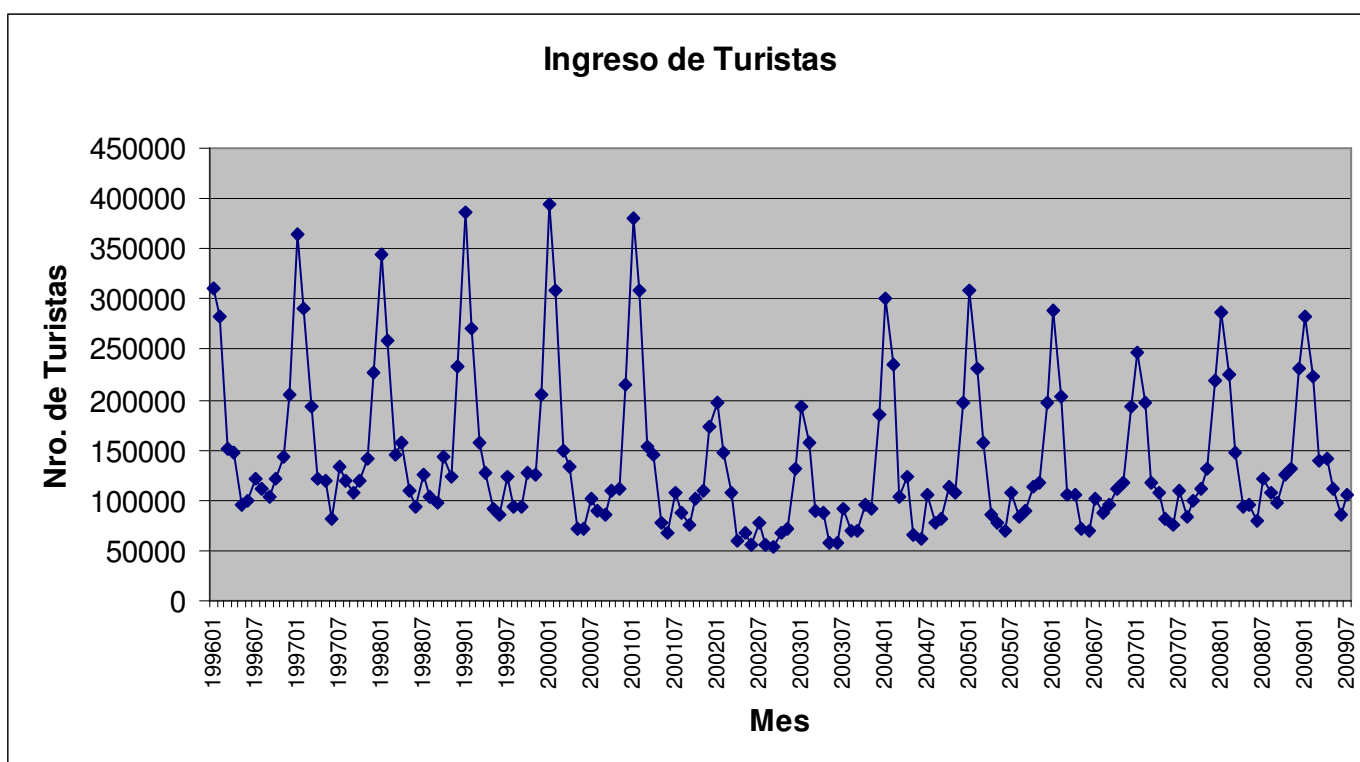


Figura 4.8. Número de turistas ingresados a Uruguay por mes en el período 1996.01-2009.07.

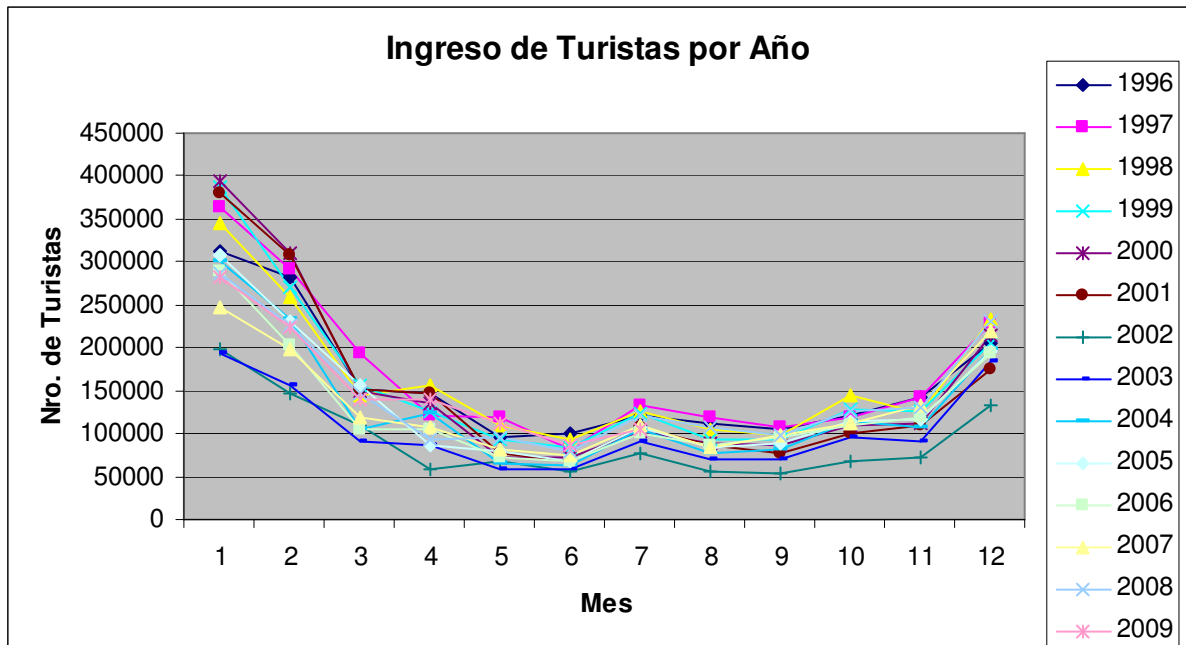


Figura 4.9. Número de turistas ingresados por año en el período 1996.01-2009.07.

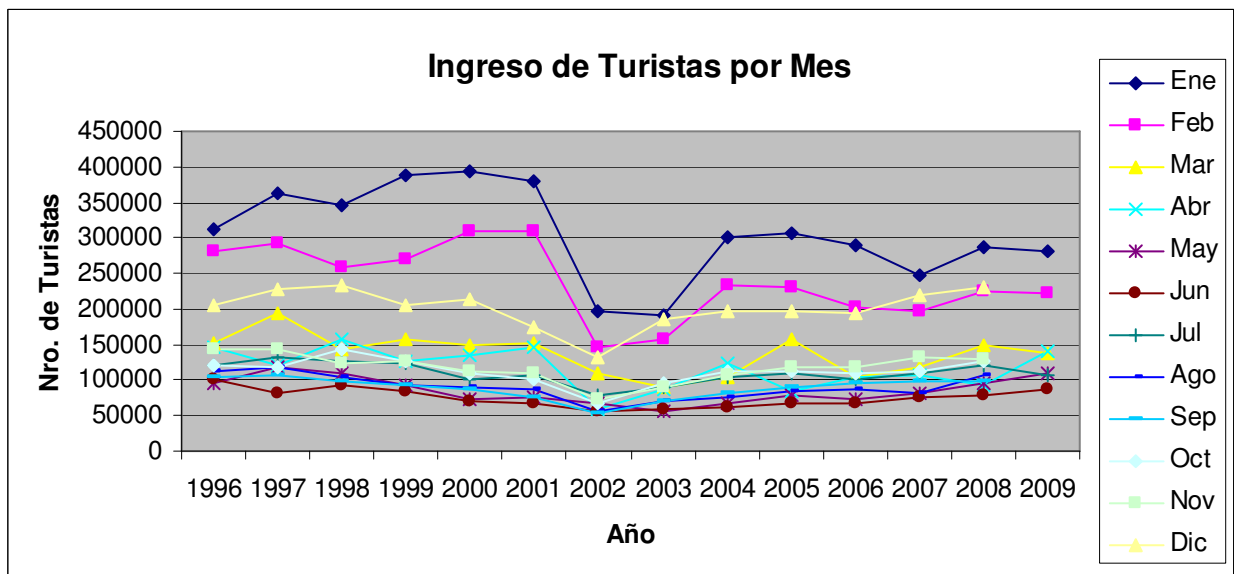


Figura 4.10. Número de turistas ingresados por mes en el período 1996.01-2009.07.

La figura 4.8 corresponde al número de turistas ingresados a Uruguay por mes. En ella se observó una fuerte estacionalidad en los meses de verano. Se observó una fuerte caída en el ingreso de turistas a partir de enero de 2002, sin volver a alcanzar nunca los niveles de los eneros anteriores a 2002.

La figura 4.8 reflejó la existencia de una serie con media y varianza no constantes lo que indicó que el proceso es no estacionario. Por tanto, la serie debió ser transformada para ser estacionaria. Para esto se aplicó logaritmo a la serie y luego se aplicó una diferencia estacional (Δ_{12}).

La figura 4.9 correspondiente al número de turistas ingresados a Uruguay por año también sugirió una fuerte estacionalidad ya que en todos los años se observó un comportamiento similar a lo largo de todos los meses; en enero ingresó el mayor número de turistas, seguido por febrero y luego por diciembre, marzo, abril, julio y luego el resto de los meses. Esto se repitió año a año.

En la figura 4.10 se corroboró lo explicado anteriormente al observar que el número de turistas ingresados al Uruguay fue mayor para los meses de enero de cada año, seguidos por los meses de febrero, luego por diciembre, marzo y abril.

En base a las figuras 4.8, 4.9 y 4.10 se pudo ver que la serie no tiene tendencia.

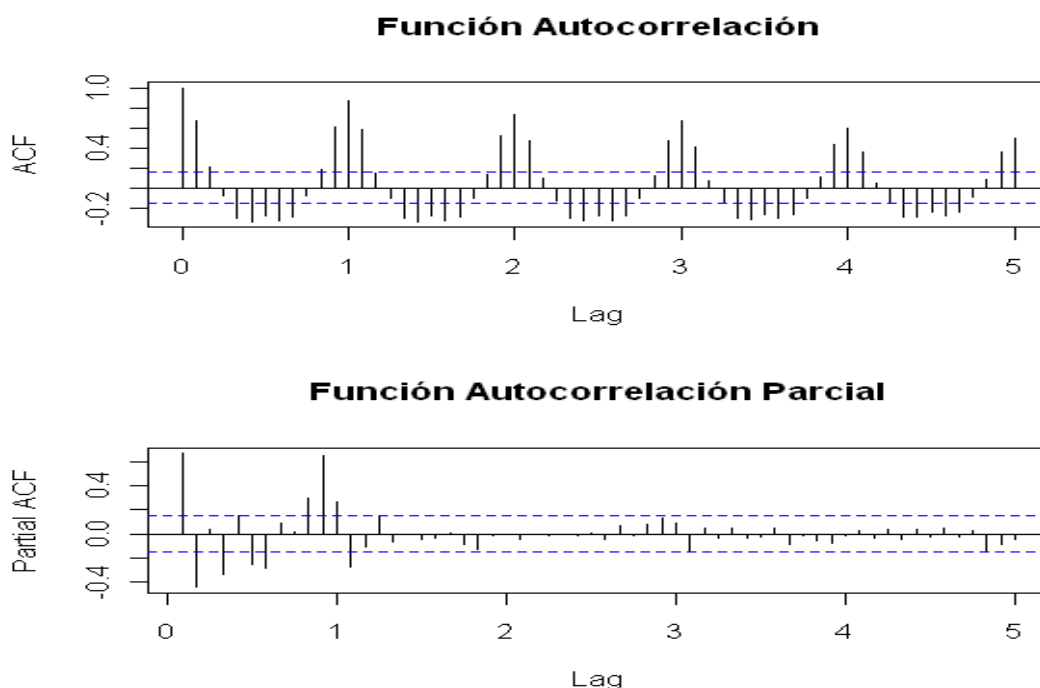


Figura 4.11. Autocorrelograma simple y parcial del Modelo 2.

En la figura 4.11 se corroboró la no estacionariedad de la serie ya que el autocorrelograma simple no decayó rápidamente, sino que tuvo un comportamiento oscilante.

En base a lo observado, para lograr estacionariedad de la serie, se decidió aplicar a la serie una diferencia estacional además de la transformación logarítmica previamente aplicada.

A continuación se presenta como quedó la serie transformada.

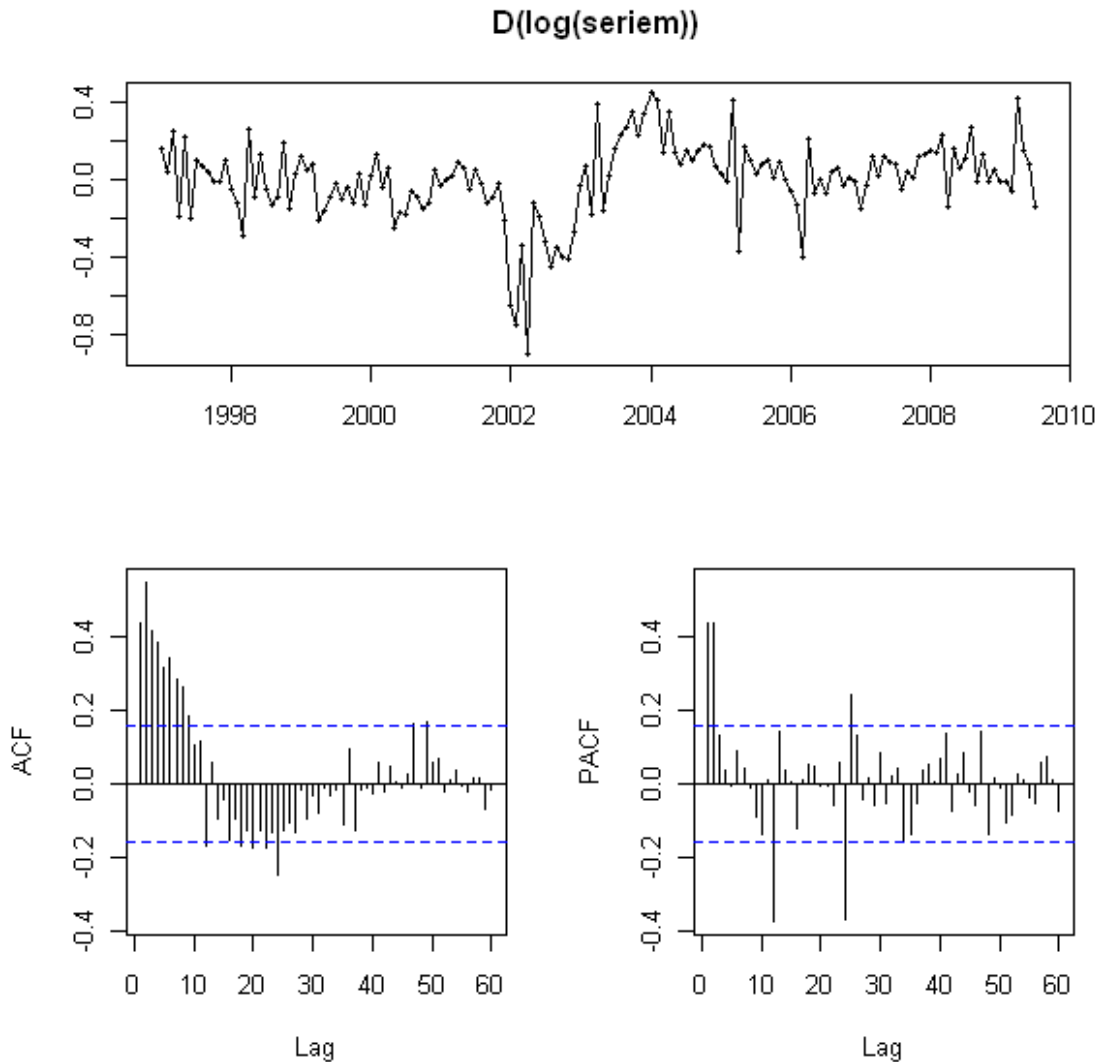


Figura 4.12. Gráfico y autocorrelograma simple y parcial de la primera diferencia estacional del logaritmo del Modelo 2 - $\Delta_{12} \log(seriem)$.

Observando la figura 4.12 se concluyó que la nueva serie transformada es estacionaria en media y varianza, ya que osciló en torno a su nivel medio. Hubo una fuerte caída en el 2002 que se debió a la incidencia de la crisis económica regional de fines del 2001 que repercutió fuertemente en el número de turistas ingresados al país a partir de enero de 2002 y en los siguientes meses.

4.2.3. Ajuste del modelo

Luego de haber analizado los gráficos de la serie original y de la serie transformada (con logaritmo y una diferencia estacional) se probó con los siguientes modelos para lograr el ajuste de los datos.

$$\text{ARIMA } (1, 0, 0) (0, 1, 1)_{(12)} \quad \text{ARIMA } (1, 0, 0) (1, 1, 0)_{(12)}$$

$$\text{ARIMA } (0, 0, 1) (0, 1, 1)_{(12)} \quad \text{ARIMA } (0, 0, 1) (1, 1, 0)_{(12)}$$

A su vez, a cada modelo se le aplicaron las siguientes intervenciones:

- **Turismo:** Se consideró la variable indicatriz Turismo para indicar en que mes del año se ubica la semana de turismo. Puede ser en marzo o en abril de cada año o en ambos.
- **Tc Enero 2002:** Se consideró la crisis del 2002 y su impacto negativo en el número de turistas que ingresaron al país a partir de enero de 2002 como un cambio transitorio.
- **Tc Mayo 2002:** Se consideró el retiro de capitales depositados en la banca uruguaya por parte de los argentinos y su impacto positivo sobre el número de turistas que ingresaron al país provenientes de Argentina a partir de mayo de 2002 como un cambio transitorio. Se supone que a partir de mayo de 2002 aumenta la cantidad de turistas argentinos que visitaron Uruguay con el único propósito de retirar capitales invertidos en la banca uruguaya ante la suposición que en Uruguay también se dispusiera un “corralito” bancario.
- **Ls Enero 2002:** Se consideró la crisis del 2002 y su impacto negativo en el número de turistas que ingresaron al país a partir de enero de 2002 como un cambio de nivel.
- **AO Mayo 2002:** También se decidió suponer que el retiro de capitales de los turistas argentinos fue un atípico aditivo. Es decir, se tomó una variable indicatriz que indique 1 en el mes mayo de 2002 y 0 en el resto de los meses.

Se consideraron las intervenciones mencionadas alternativamente de la siguiente manera:

- Turismo , Tc Enero 2002, Tc Mayo 2002
- Turismo, Tc Enero 2002

- Turismo, Ls Enero 2002, Tc Mayo 2002
- Turismo, Ls Enero 2002
- Turismo, Tc Enero 2002, AO Mayo 2002

De acuerdo a los resultados obtenidos, los siguientes modelos intermedios para el ajuste del número de turistas que ingresan a Uruguay cumplieron con los requisitos necesarios para la significación de los coeficientes estimados:

- ARIMA $(1,0,0) (0,1,1)_{(12)}$ con Turismo, Tc Enero 2002 y Tc Mayo 2002.
- ARIMA $(1,0,0) (1,1,0)_{(12)}$ con Turismo, Tc Enero 2002 y Tc Mayo 2002.
- ARIMA $(1,0,0) (0,1,1)_{(12)}$ con Turismo, Tc Enero 2002 y AO Mayo 2002.
- ARIMA $(1,0,0) (1,1,0)_{(12)}$ con Turismo, Tc Enero 2002 y AO Mayo 2002.

Se observó que la diferencia de las intervenciones es la duración del efecto de Mayo 2002. Se concluyó que para predecir el número de turistas que ingresan al país resultaron significativos el efecto de la semana de turismo y el efecto de la crisis del 2002, que se considera como cambio transitorio, es decir, su efecto va decreciendo a medida que pasa el tiempo. También resultó significativo el aluvión de turistas argentinos que ingresó al país en mayo 2002 para retirar sus capitales de los bancos, tanto como cambio transitorio y como atípico aditivo.

No resultó significativo el efecto de Enero 2002 considerado como cambio de nivel.

4.2.4. Modelo seleccionado

Se estimaron los coeficientes para estos cuatro modelos y en todos los casos obtuvimos coeficientes significativos.

Se buscó probar que intervención provocaba un menor AIC para cada ARIMA. Por lo tanto, se calculó el AIC de cada modelo con las intervenciones: Tc Mayo 2002 ó AO Mayo 2002. Los que tuvieron menor AIC fueron: ARIMA $(1,0,0) (0,1,1)_{(12)}$ y ARIMA $(1,0,0) (1,1,0)_{(12)}$ ambos con las intervenciones: Turismo, Tc Enero 2002 y Tc Mayo 2002.

Modelos	AIC
ARIMA (1,0,0) (0,1,1) ₍₁₂₎ con Turismo, Tc Enero 2002 y Tc Mayo 2002	-241,79
ARIMA (1,0,0) (0,1,1) ₍₁₂₎ con Turismo, Tc Enero 2002 y AO Mayo 2002.	-240,37

Cuadro 4.5.a. Comparación de AIC entre modelos ARIMA(1,0,0)(0,1,1)(12).

Modelos	AIC
ARIMA (1,0,0) (1,1,0) ₍₁₂₎ con Turismo, Tc Enero 2002 y Tc Mayo 2002	-232,6
ARIMA (1,0,0) (1,1,0) ₍₁₂₎ con Turismo, Tc Enero 2002 y AO Mayo 2002.	-229,95

Cuadro 4.5.b. Comparación de AIC entre modelos ARIMA(1,0,0)(1,1,0)(12).

Finalmente, también se descartó el modelo ARIMA (1,0,0) (1,1,0)₍₁₂₎ con Turismo, Tc Enero 2002 y Tc Mayo 2002¹³ ya que a la hora de realizar predicciones este modelo predijo con mayor error (mayor RECM y mayor EAM).¹⁴

El modelo seleccionado es ARIMA (1,0,0) (0,1,1)₍₁₂₎ con Turismo, Tc Enero 2002 y Tc Mayo 2002.

ARIMA (1,0,0) (0,1,1) ₍₁₂₎	Coefficientes	t observados
Ar1	0,84	18,28
Sma1	-0,59	-6,23
Turismo	0,26	12,27
Tc Enero 2002	-0,31	-3,48
Tc Mayo 2002	0,30	3,17

Cuadro 4.6. Modelo seleccionado ARIMA(1,0,0)(0,1,1)(12).

Se observó que los coeficientes del modelo estimado son significativos al 5%.

Se verificó que este modelo cumplió con los supuestos de los residuos.

¹³ El modelo descartado está desarrollado en el Apéndice D.2.

¹⁴ Ver Apéndice D.3.

El cumplimiento de los supuestos de los residuos verifica los siguientes aspectos:

- Media cero:

En la figura D.1 del apéndice D se observó que los residuos fluctúan en torno a la media que se aproxima a cero. También se observó que las barras de los coeficientes de autocorrelación de los residuos permanecieron dentro de las bandas de confianza, a excepción de los rezagos 1 y 16. Ver cuadro D.1 de las medidas de resumen de los residuos en el apéndice D.

- Incorrección:

Para testear la incorrección de los residuos se utilizó el test de Ljung-Box. Se realizó dicha prueba tomando $k = 12, 24, 36$ y 48 y resultó positiva para aceptar la hipótesis nula de ausencia de autocorrelación en los residuos, al 5%.

Se corroboró que los residuos están incorreccionados al observar que las barras de los coeficientes de autocorrelación de los residuos permanecieron dentro de las bandas de confianza, a excepción del rezago 1 y 16. Ver cuadro D.2 del test de Ljung-Box en el apéndice D.

- Normalidad:

Para testear la normalidad de los residuos se utilizó el test de Shapiro Wilk. Dicho test llevó a no rechazar la hipótesis nula de normalidad de los residuos al 5%. Ver cuadro D.3 del test de Shapiro-Wilk en el apéndice D.

4.2.5. Interpretación del modelo seleccionado

El modelo seleccionado se plantea de la siguiente manera:

$$\log(\hat{y}_t) = 0,84 \log(y_{t-1}) + \log(y_{t-12}) - 0,84 \log(y_{t-13}) + 0,59 \varepsilon_{t-12} + 0,26 \text{Turismo} - 0,31 Tc_Ene02 + 0,30 Tc_May02$$

Siendo:

- Turismo la indicatriz que toma valor 1 en los meses del año en que cae la semana de Turismo y 0 en el resto.

$$I_t^{Turismo} = \begin{cases} 1, t = Turismo \\ 0, t \neq Turismo \end{cases}$$

- Tc_Ene02 el cambio transitorio de la crisis económica regional que comenzó en Diciembre 2001 y que afectó a la baja el ingreso de turistas a Uruguay a partir de enero de 2002 en adelante. Como su efecto es transitorio, el mismo va disminuyendo a medida que transcurre el tiempo hasta desaparecer. Toma valor 1 en enero de 2002 y luego disminuye de la siguiente manera:

$$\frac{1}{1-0,7L} = (1 + 0,7 + 0,7^2 + 0,7^3 + \dots)$$

- Tc_May02 el cambio transitorio debido al retiro de capitales depositados en la banca uruguaya por parte de los argentinos que afectó positivamente sobre el número de turistas que ingresaron al país provenientes de Argentina a partir de mayo de 2002 en adelante. Como su efecto es transitorio, el mismo va disminuyendo a medida que transcurre el tiempo hasta desaparecer.

La estimación y predicción en logaritmos en el momento t depende de:

- La observación del mes anterior en logaritmos $\log(y_{t-1})$, cuyo peso en la generación de la nueva observación es de 0,84.
- La observación en logaritmos de doce meses atrás $\log(y_{t-12})$, con peso 1.
- La observación en logaritmos de trece meses atrás $\log(y_{t-13})$, con peso -0,84.
- El error de predicción de doce meses atrás ε_{t-12} , con peso 0,59.

Es decir, $\log(\hat{y}_t)$ depende de la observación del mes anterior, la observación del mismo mes correspondiente al año anterior, la observación del año anterior más un mes y el error de predicción del año anterior. Cada uno con sus respectivos pesos.

A esto se le agrega el efecto del análisis de intervención:

- Turismo es una variable indicatriz que toma valor uno en los meses en que cae la semana de turismo y cero en el resto de los meses. Tiene peso 0,26.

- Tc_Ene02 es el cambio transitorio que considera a la crisis económica del 2002 y su efecto en el número de turistas que ingresan al país es -0,31. Como ya ha sido explicado, al ser un cambio transitorio el efecto de la crisis del 2002 va disminuyendo a medida que transcurre el tiempo.
- Tc_May02 es el cambio transitorio que considera al retiro de capitales depositados en la banca uruguaya por parte de los argentinos y su efecto en el número de turistas que ingresan al país es 0,30. Como su efecto es transitorio, el mismo va disminuyendo a medida que transcurre el tiempo hasta desaparecer.

Se interpretó que la predicción del número de turistas que ingresará a Uruguay depende del número de turistas que haya ingresado un mes y un año atrás con efecto positivo y se notó que incide más el número de turistas que ingresaron un año atrás que un mes atrás debido a que esta serie es claramente estacional. El número de turistas que ingresaron trece meses atrás tiene un efecto negativo en la predicción del número de turistas que ingresará a Uruguay. Además depende del error de predicción del año anterior.

También se interpretó que el mes en que cae la semana de turismo y el cambio transitorio de Mayo 2002 tienen efecto positivo y la crisis del 2002 tiene efecto negativo decreciente en la predicción del número de turistas que ingresará a Uruguay.

4.2.6. Predicción

4.2.6.1. Predicción dentro de la muestra

Para probar el modelo estimado se cortó la serie y se generaron predicciones dentro de la muestra para dicho intervalo.

Mes	Predicción	Valor Real	Int. Inf. 95	Int. Sup. 95	Error	Error en %
Ago-08	93.640	108.215	76.821	114.141	14.575	13,5%
Set-08	101.380	97.549	78.147	131.520	-3.831	-3,9%
Oct-08	120.038	126.139	89.143	161.640	6.101	4,8%
Nov-08	128.736	130.551	93.294	177.644	1.815	1,4%
Dic-08	216.024	231.510	153.957	303.113	15.486	6,7%
Ene-09	295.205	281.925	207.948	419.077	-13.280	-4,7%
Feb-09	225.700	223.561	157.679	323.064	-2.139	-1,0%
Mar-09	117.525	139.164	81.623	169.219	21.639	15,5%
Abr-09	117.669	140.922	81.378	170.146	23.253	16,5%
May-09	84.979	110.788	58.591	123.254	25.809	23,3%
Jun-09	75.393	86.307	51.866	109.591	10.914	12,6%
Jul-09	113.719	106.016	78.107	165.568	-7.703	-7,3%
Total	1.690.009	1.782.647			92.638	5,2%

Cuadro 4.7. Predicción dentro de la muestra a 12 pasos para el modelo seleccionado ARIMA(1,0,0)(0,1,1)(12).

Observando el cuadro 4.7 se concluyó que los valores reales de la serie estaban dentro del intervalo de confianza para los 12 períodos.

Para cada caso, se comparó el valor real observado con la predicción del modelo calculando el error y luego el error relativo porcentual de cada predicción.

Si bien hay meses en los cuales se encontraron errores altos (por ejemplo Mayo 2009 con 23,3%), dentro del año el modelo tuvo un error relativo porcentual de 5,2%.

En particular, si se consideran los meses de verano diciembre, enero y febrero en total presentan un error de 0,01%.

Mes	Predicción	Valor Real	Int. Inf. 95	Int. Sup. 95	Error	Error en %
Dic-08	216.024	231.510	153.957	303.113	15.486	6,7%
Ene-09	295.205	281.925	207.948	419.077	-13.280	-4,7%
Feb-09	225.700	223.561	157.679	323.064	-2.139	-1,0%
Total	736.929	736.996			0.067	0,01%

Cuadro 4.8. Predicción dentro de la muestra para diciembre 2008, enero y febrero 2009.

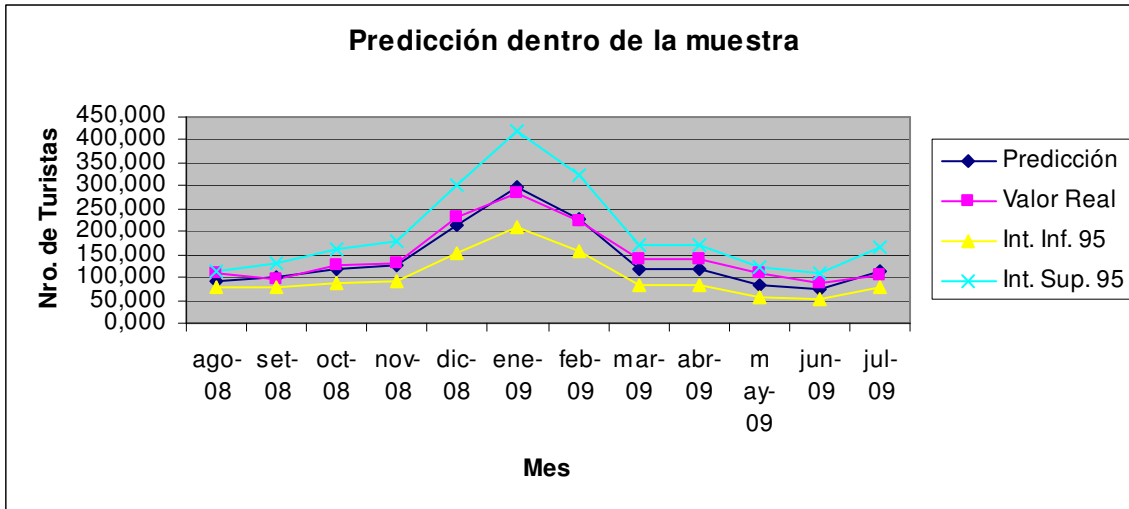


Figura 4.13. Predicción dentro de la muestra a 12 pasos para el modelo seleccionado ARIMA(1,0,0)(0,1,1)(12).

Posteriormente se realizaron las predicciones dentro de la muestra para el modelo seleccionado actualizando a un paso, es decir incluyendo un valor observado real con el objetivo de ver si las predicciones mejoraron al incorporar observaciones reales.

Observando los cuadros D.6 hasta D.16 del apéndice D se concluyó que los valores reales de la serie están dentro del intervalo de confianza al 95% para todos los períodos, lo cual indicó que las predicciones realizadas por el modelo seleccionado son correctas.

Para cada caso, se comparó el valor real observado con la predicción del modelo calculando el error y luego el error porcentual de cada predicción.

Se observó que a medida que se incorporaron observaciones reales al modelo estimado, las predicciones mejoraron, ya que los errores fueron disminuyendo.

4.2.6.2. Predicción fuera de la muestra

Mes	Predicción	Int. Inf. 95	Int. Sup. 95
Ago-09	87.906	71.944	107.409
Set-09	90.205	69.416	117.219
Oct-09	112.234	83.310	151.199
Nov-09	120.399	87.328	165.995
Dic-09	208.739	149.083	292.267
Ene-10	274.442	193.955	388.329
Feb-10	214.727	150.653	306.054
Mar-10	158.676	110.763	227.313
Abr-10	122.269	85.047	175.782
May-10	92.138	63.929	132.795
Jun-10	77.857	53.925	112.410
Jul-10	108.518	75.068	156.874

Cuadro 4.9. Predicción fuera de la muestra para el modelo seleccionado ARIMA(1,0,0)(0,1,1)(12).

Predicción fuera de la muestra

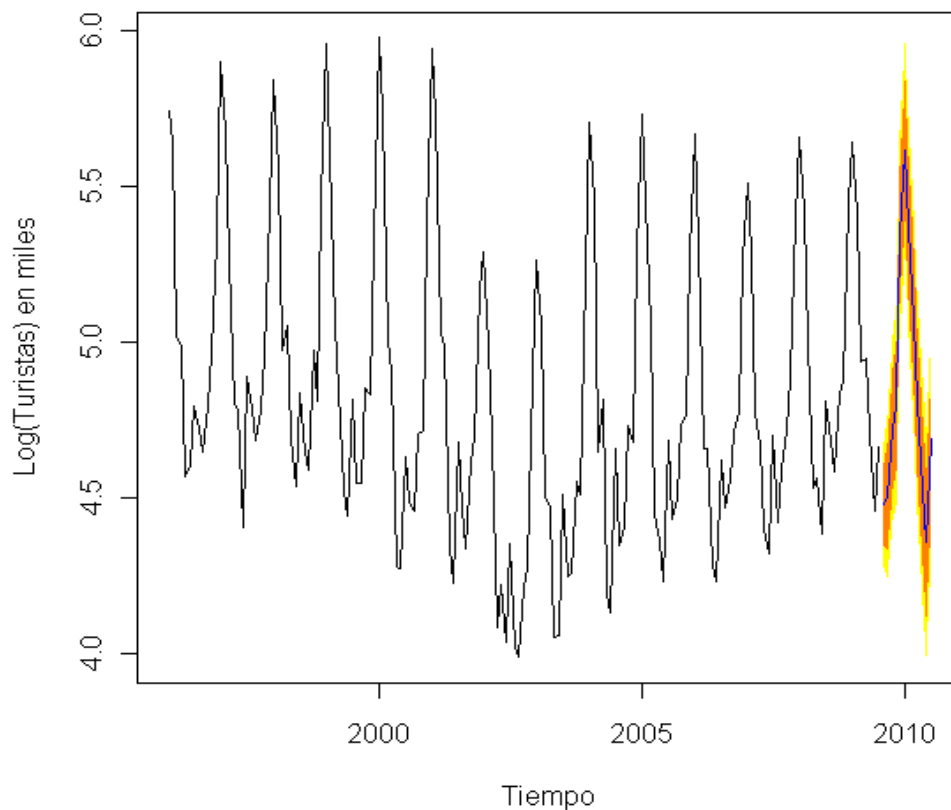


Figura 4.14. Predicción fuera de la muestra para el modelo seleccionado ARIMA(1,0,0)(0,1,1)(12).

Mes	2008/2009	2009/2010	Tasa de crecimiento
Agosto	108.215	87.906	-18,77%
Setiembre	97.549	90.205	-7,53%
Octubre	126.139	112.234	-11,02%
Noviembre	130.551	120.399	-7,78%
Diciembre	231.510	208.739	-9,84%
Enero	281.925	274.442	-2,65%
Febrero	223.561	214.727	-3,95%
Marzo	139.164	158.676	14,02%
Abril	140.922	122.269	-13,24%
Mayo	110.788	92.138	-16,83%
Junio	86.307	77.857	-9,79%
Julio	106.016	108.518	2,36%

Cuadro 4.10. Comparación con igual mes del año anterior.

Observando el cuadro 4.10 se pudo pronosticar que es de esperar que el número de turistas que ingresará a Uruguay en los próximos doce meses sea inferior a la cantidad de turistas ingresados en los últimos doce meses, con excepción de los meses marzo y julio.

Esto podría explicarse por la crisis financiera internacional que estalló en EEUU en octubre del 2008 pero que tuvo un efecto retardado en el poder adquisitivo de los turistas que ingresarán a Uruguay de Agosto 2009 a Julio 2010.

4.3. Comparación de los modelos 1 y 2

Se concluyó que si se considera el modelo 2 las predicciones indicarían que el número de turistas que ingresarán a Uruguay en los próximos doce meses disminuirá 6,42%, según cuadro 4.10, mientras que si se considera el modelo 1 las predicciones indicarían que el número de turistas que ingresarán a Uruguay en los próximos cuatro trimestres aumentará 6,24%, según cuadro 4.4. Se recuerda que la serie mensual no incluye a los turistas uruguayos residentes en el exterior mientras que la serie trimestral si los incluye. Se entendió que la explicación puede estar en parte dada por esa diferencia.

4.4. Modelización de la diferencia de la serie trimestral 1993.01 – 2009.01 y la serie mensual 1996.01 – 2009.07 – Modelo 3

4.4.1. Datos

El modelo que se presenta a continuación se construyó como diferencia algebraica entre el Modelo 1 y el Modelo 2 con frecuencia trimestral en lugar de mensual. Se generó una nueva serie trimestral que abarca el período 1996.01 a 2009.01, llamada Modelo 3.

Esta serie incluye a los extranjeros que perciben algún tipo de remuneración en Uruguay y a los turistas uruguayos residentes en el exterior que ingresan a Uruguay, mientras que el Modelo 2 no los incluye. Dado que los extranjeros que perciben una remuneración en Uruguay son muy pocos, se supuso que el nuevo modelo se compone mayoritariamente de los turistas uruguayos residentes en el exterior.

Se encontró de interés analizar el comportamiento de los turistas uruguayos residentes en el exterior ya que en un análisis preliminar del modelo se observó que el mismo no tiene el mismo comportamiento que los dos modelos analizados anteriormente.

4.4.2. Análisis descriptivo de la serie

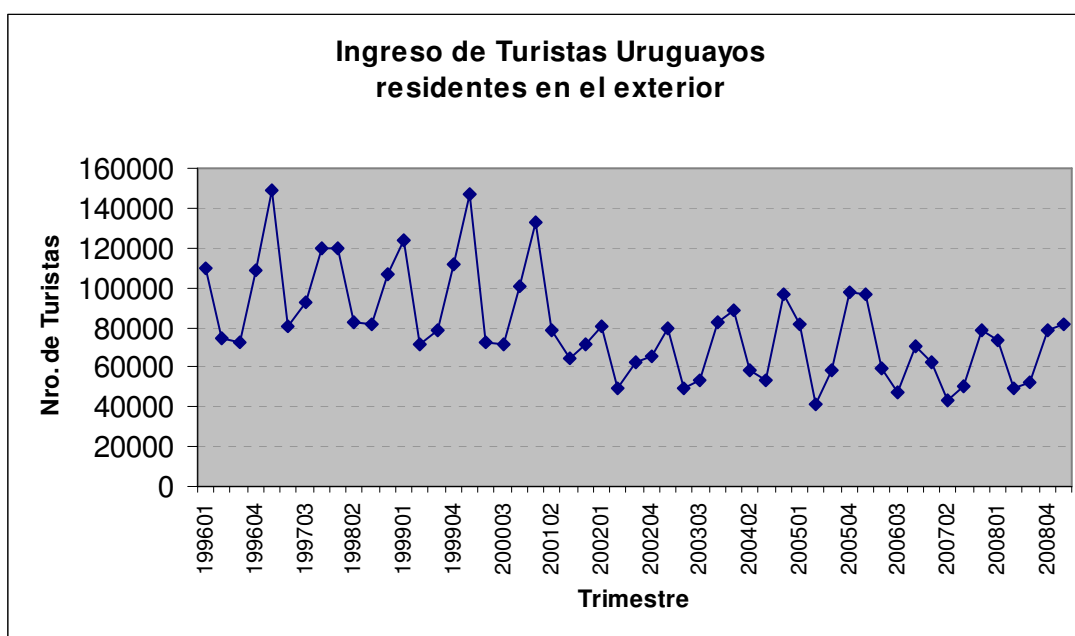


Figura 4.15. Número de turistas uruguayos residentes en el exterior ingresados a Uruguay por trimestre en el período 1996.I–2009.I.

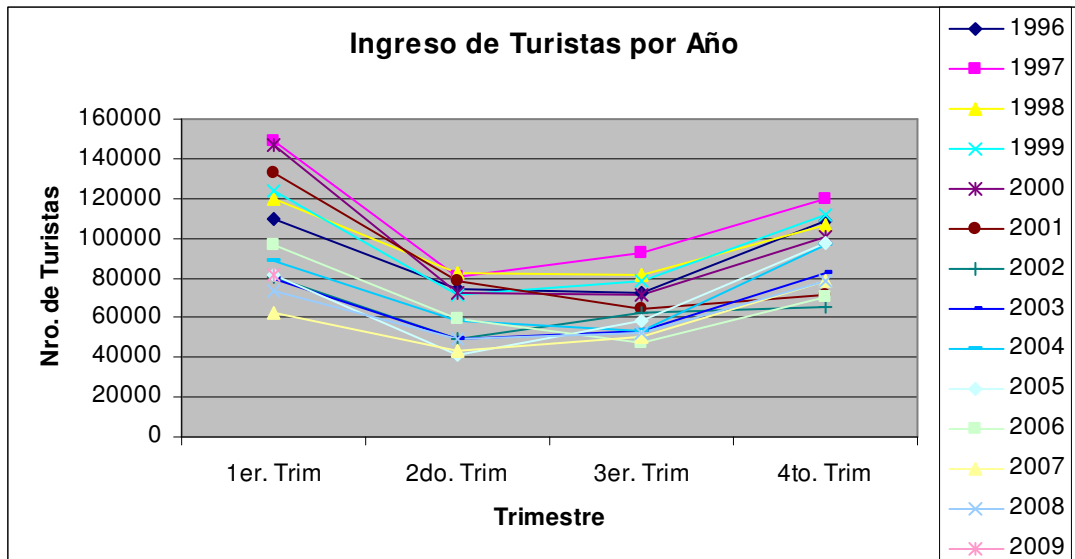


Figura 4.16. Número de turistas uruguayos residentes en el exterior ingresados por año en el período 1996.I–2009.I.

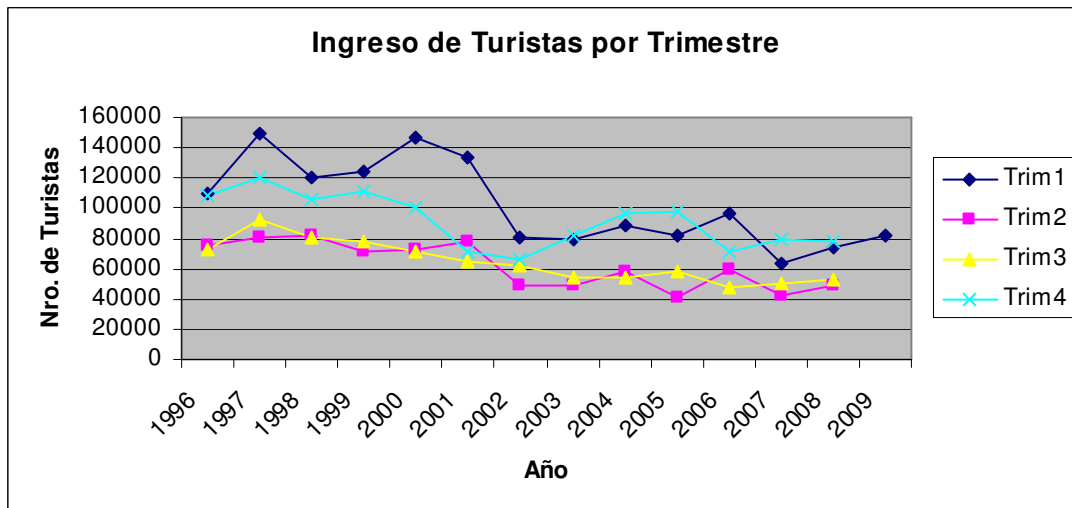


Figura 4.17. Número de turistas uruguayos residentes en el exterior ingresados por trimestre en el período 1996.I–2009.I.

La figura 4.15 corresponde al número de turistas uruguayos residentes en el exterior ingresados a Uruguay por trimestre. En ella se observó, al igual que en las series anteriores, una fuerte estacionalidad en los meses de verano. Además se observó una fuerte caída en el ingreso de turistas uruguayos residentes en el exterior a partir del primer trimestre del 2002, sin volver a alcanzar nunca los niveles de los trimestres anteriores. Como algo particular de esta serie se notó que el valor más bajo de turistas uruguayos residentes en el exterior que recibió Uruguay no se dio en el 2002, sino que ocurrió en el 2005. Es por este motivo que se supuso que la mayoría de los turistas uruguayos residentes en el exterior provienen de Argentina, ya que se pensó en el impacto del corte de los puentes entre Argentina y Uruguay desde Diciembre 2005, a

partir del cual se produjo la disminución en el número de turistas uruguayos residentes en el exterior ingresados a Uruguay.

La figura 4.15 reflejó la existencia de una serie con media y varianza no constantes lo que indicó que el proceso es no estacionario. Por tanto, la serie debió ser transformada para ser estacionaria. Para esto se aplicó logaritmo a la serie y luego se aplicó una diferencia estacional (Δ_4).

Las figuras 4.16 y 4.17 corresponden al número de turistas uruguayos residentes en el exterior ingresados a Uruguay por año. Ambas sugirieron una fuerte estacionalidad ya que en todos los años se observó un comportamiento similar a lo largo de los trimestres; en el primer trimestre ingresa el mayor número de turistas, seguido por el cuarto trimestre y luego alternan entre el segundo y tercer trimestre.

Hubo un posible cambio en la estacionalidad ya que se notó que esto se revirtió a partir del 2003 ya que luego de este año se observó que ingresaron más turistas uruguayos residentes en el exterior en el cuarto trimestre que en el primer trimestre. En el 2006 se volvieron a recibir más turistas uruguayos residentes en el exterior en el primer trimestre que en el cuarto trimestre. Se notó entonces que el comportamiento de este segmento de turistas es diferente al resto de los turistas, cuyo número es siempre mayor en el primer trimestre, luego le sigue en cantidad el cuarto trimestre y posteriormente alternan entre el segundo y tercer trimestre.

También se observó un cambio de nivel desde el año 2002 ya que nunca se logró alcanzar los niveles anteriores de ingreso de turistas. Por otro lado, los modelos 1 y 2 se recuperaron mejor de la crisis del 2002, aunque tampoco llegaron a alcanzar los valores anteriores de ingreso de turistas. Esto se confirmó con las distintas intervenciones que se le aplicaron a los dos modelos anteriores. En el modelo 1 las intervenciones fueron Turismo y Tc2002 y en el modelo 2 las intervenciones fueron Turismo, Tc Enero 2002 y Tc Mayo 2002. Por esto se supuso que el modelo 3 será intervenido con Ls2002.

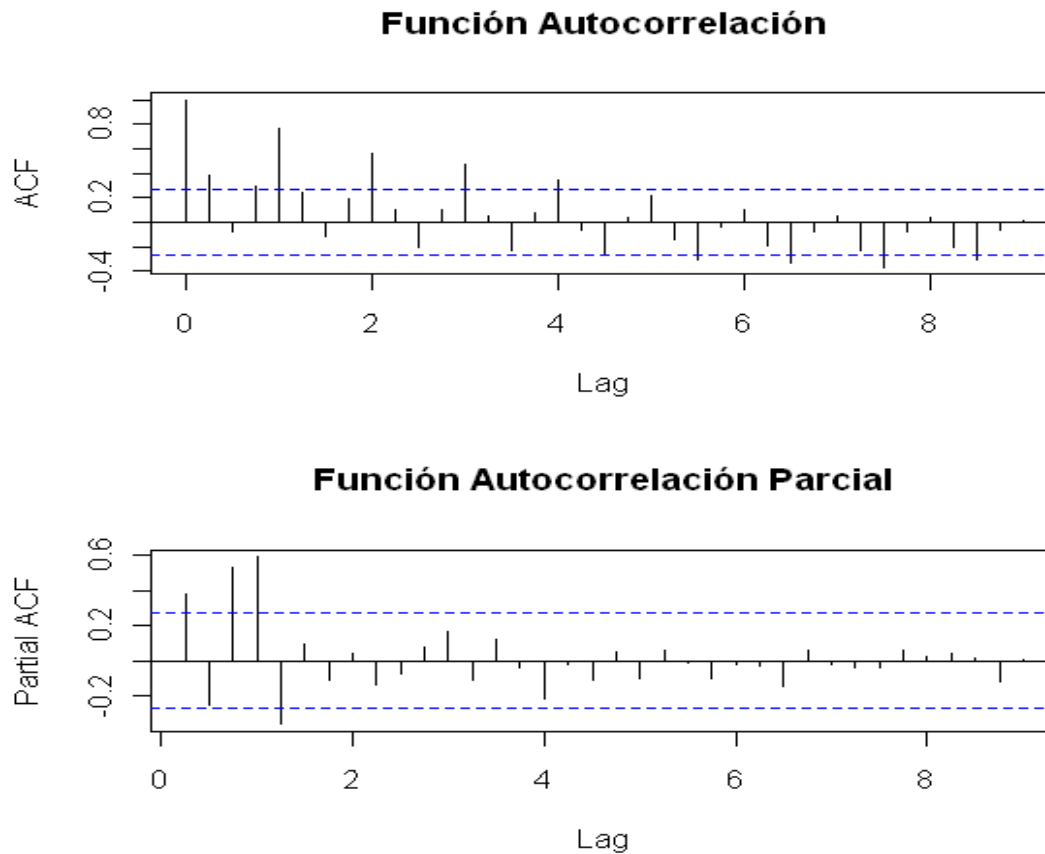


Figura 4.18. Autocorrelograma simple y parcial del Modelo 3.

En base a lo observado en la figura 4.18, para lograr estacionariedad de la serie, se decidió aplicar a la serie una diferencia estacional además de la transformación logarítmica previamente aplicada.

A continuación se presenta como quedó la serie transformada.

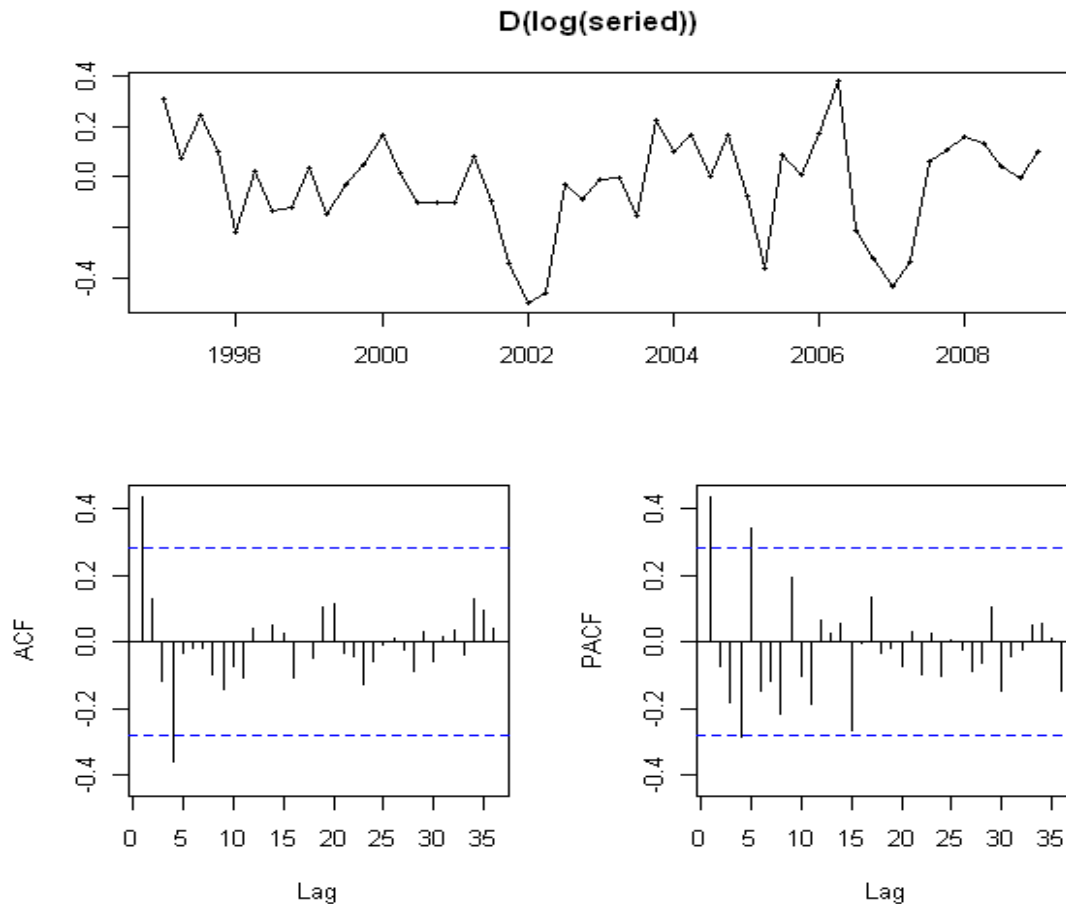


Figura 4.19. Gráfico y autocorrelograma simple y parcial de la primera diferencia estacional del logaritmo del Modelo 3 - $\Delta_4 \log(\text{serie})$.

Observando la figura 4.19 se concluyó que la nueva serie transformada es estacionaria en media y varianza, ya que osciló en torno a su nivel medio. Hubo una fuerte caída en el 2002 que se debió a la incidencia de la crisis económica regional de fines del 2001 que repercutió fuertemente en el número de turistas uruguayos residentes en el exterior ingresados al país a partir del primer trimestre 2002 y los siguientes trimestres. También se observó una fuerte caída en el ingreso de turistas uruguayos residentes en el exterior del año 2006 al 2007 debido al corte de los puentes.

4.4.3. Ajuste del modelo

Luego de haber analizado los gráficos de la serie original y de la serie transformada (con logaritmo y una diferencia estacional) se probó con los siguientes modelos para lograr el ajuste de los datos.

ARIMA (1,0,0) (0,1,1)₍₄₎

ARIMA (1,0,0) (1,1,0)₍₄₎

ARIMA (0,0,1) (0,1,1)₍₄₎

ARIMA (0,0,1) (1,1,0)₍₄₎

A su vez, a cada modelo se le aplicaron las siguientes intervenciones:

- **Turismo:** Se consideró la variable indicatriz Turismo para indicar en que trimestre del año se ubicó la semana de turismo. Puede ser en el primer o segundo trimestre de cada año o en ambos.
- **Tc2002:** Se consideró como cambio transitorio la crisis del 2002 y su efecto en el número de turistas uruguayos residentes en el exterior que ingresaron al país.
- **Tc2006:** Se consideró como cambio transitorio al impacto debido al corte de los puentes entre Argentina y Uruguay que comenzó en Diciembre 2005 y afectó el número de turistas uruguayos residentes en el exterior que ingresaron al país provenientes de Argentina a partir del primer trimestre del 2006.
- **Ls2002:** Se consideró como cambio de nivel la crisis del 2002 y su efecto en el número de turistas uruguayos residentes en el exterior que ingresaron al país.
- **Ls2006:** Se consideró como cambio de nivel al impacto debido al corte de los puentes entre Argentina y Uruguay que comenzó en Diciembre 2005 y afectó el número de turistas uruguayos residentes en el exterior que ingresaron al país provenientes de Argentina a partir del primer trimestre del 2006.

Se consideraron las intervenciones mencionadas alternativamente de la siguiente manera:

- Turismo , Tc 2002, Tc 2006
- Turismo, Ls 2002, Ls 2006
- Turismo, Tc 2002
- Turismo, Ls 2002
- Tc 2002
- Ls 2002

De acuerdo a los resultados obtenidos, los siguientes modelos intermedios para el ajuste del número de turistas uruguayos residentes en el exterior que ingresan a Uruguay cumplieron con los requisitos necesarios para la significación de los coeficientes estimados:

- ARIMA (1,0,0) (0,1,1)₍₄₎ con Turismo y Ls2002
- ARIMA (1,0,0) (1,1,0)₍₄₎ con Turismo y Ls2002
- ARIMA (1,0,0) (0,1,1)₍₄₎ con Ls2002
- ARIMA (1,0,0) (1,1,0)₍₄₎ con Ls2002
- ARIMA (0,0,1) (0,1,1)₍₄₎ con Turismo y Ls2002
- ARIMA (0,0,1) (1,1,0)₍₄₎ con Turismo y Ls2002
- ARIMA (0,0,1) (0,1,1)₍₄₎ con Ls2002
- ARIMA (0,0,1) (1,1,0)₍₄₎ con Ls2002

Se concluyó que para predecir el número de turistas uruguayos residentes en el exterior que ingresan al país resultaron significativos los coeficientes que tuvieron en cuenta el efecto de la semana de turismo y el efecto de la crisis del 2002, que se consideró como un cambio de nivel, es decir, su efecto se mantiene hasta el final de la serie analizada.

No resultaron significativos para nuestro análisis el efecto del corte de los puentes del 2006, ya sea considerado como cambio transitorio o de nivel y el efecto de la crisis del 2002 considerado como cambio transitorio.

4.4.4. Modelo seleccionado

Se estimaron los coeficientes para estos ocho modelos y en todos los casos se obtuvieron coeficientes significativos.

Para seleccionar el modelo estimado se realizó una tabla comparando el AIC y el RECM de estos ocho modelos y se seleccionó el modelo con menor RECM y con AIC pequeño.¹⁵

¹⁵ El modelo descartado se encuentra desarrollado en el Apéndice E.2.

Modelos	AIC	RECM (predicción dentro de la muestra a 4 pasos)
ARIMA (0,0,1) (0,1,1) ₍₄₎ con Ls2002	-45,53	1,39
ARIMA (1,0,0) (0,1,1) ₍₄₎ con Ls2002	-45,79	1,43
ARIMA (1,0,0) (0,1,1) ₍₄₎ con Turismo y Ls2002	-46,47	4,96
ARIMA (0,0,1) (0,1,1) ₍₄₎ con Turismo y Ls2002	-46,23	5,20
ARIMA (0,0,1) (1,1,0) ₍₄₎ con Ls2002	-38,71	7,33
ARIMA (1,0,0) (1,1,0) ₍₄₎ con Ls2002	-38,51	7,44
ARIMA (0,0,1) (1,1,0) ₍₄₎ con Turismo y Ls2002	-40,73	8,88
ARIMA (1,0,0) (1,1,0) ₍₄₎ con Turismo y Ls2002	-39,48	9,02

Cuadro 4.11. Comparación de RECM y AIC entre los ocho modelos.

El modelo seleccionado es ARIMA (0,0,1) (0,1,1)₍₄₎ con Ls2002.

Se concluyó que para predecir el número de turistas uruguayos residentes en el exterior que ingresan al país no resultó significativo el efecto de la semana de turismo. Es esperable, pues se entiende que los turistas uruguayos residentes en el exterior no toman como determinante la semana de turismo para viajar a Uruguay, como si lo pueden ser las fiestas en diciembre.

ARIMA (0,0,1) (0,1,1) ₍₄₎	Coefficientes	t observados
Ma1	0,32	2,51
Sma1	-0,77	-4,96
Ls2002	-0,33	-4,88

Cuadro 4.12. Modelo seleccionado ARIMA(0,0,1)(0,1,1)(4).

Se observó que los coeficientes del modelo estimado son significativos al 5%.

Se verificó que este modelo cumplió con los supuestos de los residuos.

El cumplimiento de los supuestos de los residuos verifica los siguientes aspectos:

- Media cero:

En la figura E.1 del apéndice E se observó que los residuos fluctúan en torno a la media que se aproxima a cero. Además se observó que las barras de los coeficientes de autocorrelación de los residuos permanecieron dentro de las bandas de confianza. Ver cuadro E.1 de las medidas de resumen de los residuos en el apéndice E.

- Incorrelación:

Para testear la incorrelación de los residuos se utilizó el test de Ljung-Box. La prueba se realizó tomando $k = 4, 8, 12$ y 16 y resultó positiva para aceptar la hipótesis nula de ausencia de autocorrelación en los residuos, al 5%.

Se corroboró que los residuos están incorrelacionados al observar que las barras de los coeficientes de autocorrelación de los residuos permanecieron dentro de las bandas de confianza, evidenciando que la dinámica de la serie fue captada por la modelización propuesta. Ver cuadro E.2 del test de Ljung-Box en el apéndice E.

- Normalidad:

Para testear la normalidad de los residuos se utilizó el test de Shapiro Wilk y el test de Jarque Bera. Los mismos llevaron a no rechazar la hipótesis nula de normalidad de los residuos, al 5%. Ver cuadro E.3 del test de Shapiro-Wilk y Jarque-Bera en el apéndice E.

4.4.5. Interpretación del modelo seleccionado

El modelo seleccionado se plantea de la siguiente manera:

$$\log(\hat{y}_t) = \log(y_{t-4}) - 0,32\varepsilon_{t-1} + 0,77\varepsilon_{t-4} - 0,25\varepsilon_{t-5} - 0,33Ls2002$$

Siendo:

- Ls2002 el cambio de nivel de la crisis económica regional que comenzó en Diciembre 2001 y que afectó a la baja el ingreso de turistas uruguayos residentes en el exterior a Uruguay a partir del primer trimestre del año 2002 en

adelante. Como su efecto es permanente el mismo se mantiene hasta el final de la serie. Toma valor 1 a partir del primer trimestre 2002 hasta el final de la serie.

$$Ls2002 = S_t^{1T2002} = \begin{cases} 0, t < 1T2002 \\ 1, t \geq 1T2002 \end{cases}$$

La estimación y predicción en logaritmos en el momento t depende de:

- La observación en logaritmos de cuatro trimestres atrás $\log(y_{t-4})$, con peso 1.
- El error de predicción de un trimestre atrás ε_{t-1} , con peso -0,32.
- El error de predicción de cuatro trimestres atrás ε_{t-4} , con peso 0,77.
- El error de predicción de cinco trimestres atrás ε_{t-5} , con peso -0,25.

Es decir, $\log(\hat{y}_t)$ depende de la observación del mismo trimestre correspondiente al año anterior, el error de predicción de un trimestre atrás, el error de predicción del año anterior y el error de predicción del año anterior más un trimestre. Cada uno con sus respectivos pesos.

A esto se le agrega el efecto del análisis de intervención:

- Ls2002 es el cambio de nivel que considera a la crisis económica del 2002 y su efecto en el número de turistas uruguayos residentes en el exterior que ingresan al país es -0,3331. Como ya ha sido explicado, al ser un cambio de nivel el efecto de la crisis del 2002 permanece en el tiempo hasta el final de la serie.

Se interpretó que la predicción del número de turistas uruguayos residentes en el exterior que ingresará a Uruguay depende del número de turistas uruguayos residentes en el exterior que haya ingresado un año atrás con efecto positivo y de los errores de predicción en el trimestre anterior, en el año anterior y en cinco trimestres atrás, con sus respectivos pesos.

También se interpretó que la crisis del 2002 tiene un efecto permanente en la predicción del número de turistas uruguayos residentes en el exterior que ingresará a Uruguay.

4.4.6. Predicción

4.4.6.1. Predicción dentro de la muestra

Para probar el modelo estimado se cortó la serie y se generaron predicciones dentro de la muestra para dicho intervalo.

Trimestre	Predicción	Valor Real	Int. Inf. 95	Int. Sup. 95	Error	Error en %
2T-2008	48.191	48.938	36.532	63.572	0.747	1,5%
3T-2008	52.606	52.516	39.282	70.451	-0.090	-0,2%
4T-2008	79.187	78.588	59.130	106.048	-0.599	-0,8%
1T-2009	79.164	81.791	59.114	106.014	2.627	3,2%

Cuadro 4.13. Predicción dentro de la muestra a 4 pasos para el modelo seleccionado ARIMA(0,0,1)(0,1,1)(4).

Observando el cuadro 4.13 se concluyó que los valores reales de la serie estaban dentro del intervalo de confianza para los 4 períodos.

Para cada caso, se comparó el valor real observado con la predicción del modelo calculando el error y luego el error relativo porcentual de cada predicción.

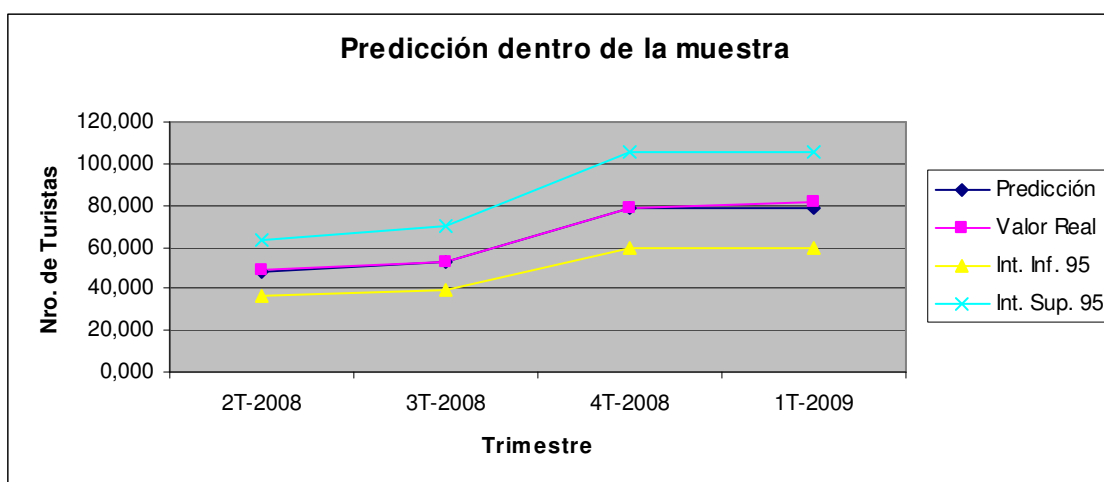


Figura 4.20. Predicción dentro de la muestra a 4 pasos para el modelo seleccionado ARIMA(0,0,1)(0,1,1)(4).

Posteriormente se realizaron las predicciones dentro de la muestra para el modelo seleccionado actualizando a un paso, es decir incluyendo un valor observado real con el objetivo de ver si las predicciones mejoraban al incorporar observaciones reales.

Observando los cuadros E.5, E.6, E.7 y E.8 del apéndice se concluyó que los valores reales de la serie están dentro del intervalo de confianza al 95% para todos los períodos, lo cual indicó que las predicciones realizadas por el modelo seleccionado fueron correctas.

Para cada caso, se comparó el valor real observado con la predicción del modelo calculando el error y luego el error porcentual de cada predicción.

Se observó que a medida que se incorporaron observaciones reales al modelo estimado, las predicciones mejoraban, ya que los errores fueron disminuyendo.

4.4.6.2. Predicción fuera de la muestra

Trimestre	Predicción	Int. Inf. 95	Int. Sup. 95
2T-2009	50.169	38.514	65.351
3T-2009	52.754	39.951	69.661
4T-2009	78.740	59.630	103.974
1T-2010	80.411	60.897	106.178

Cuadro 4.14. Predicción fuera de la muestra para el modelo seleccionado ARIMA(0,0,1)(0,1,1)(4).

Predicción fuera de la muestra

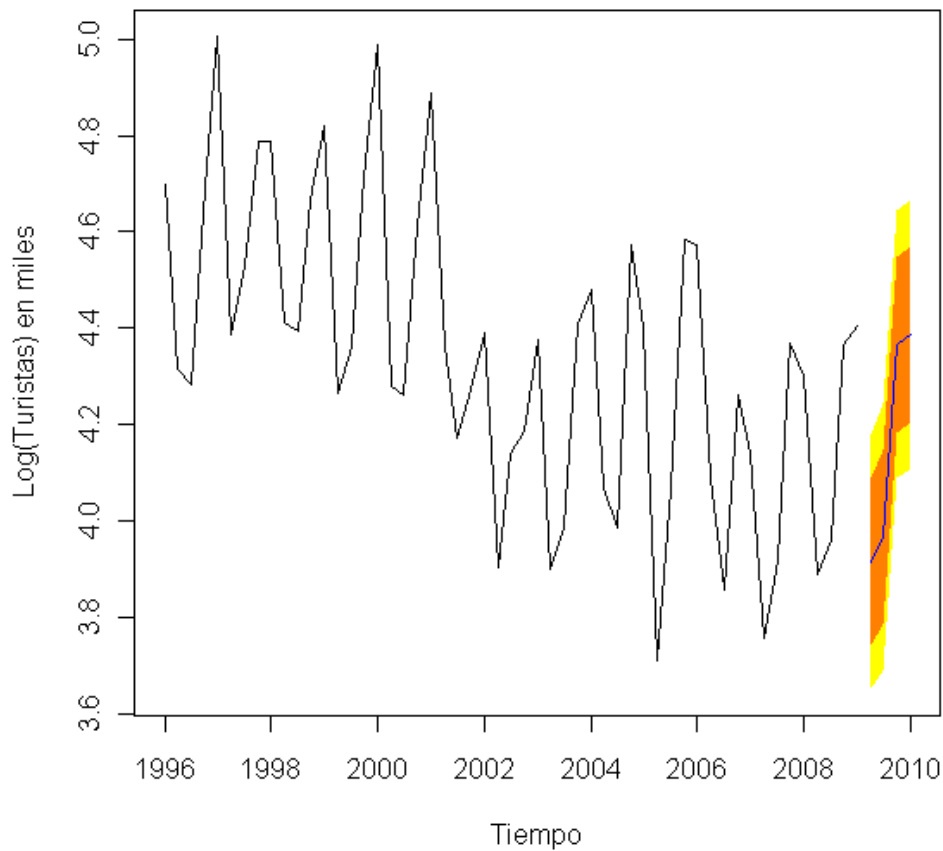


Figura 4.21. Predicción fuera de la muestra para el modelo seleccionado $ARIMA(0,0,1)(0,1,1)(4)$.

Trimestre	2008/2009	2009/2010	Tasa de Crecimiento
2T	48.938	50.169	2,52%
3T	52.516	52.754	0,45%
4T	78.588	78.740	0,19%
1T	81.791	80.411	-1,69%

Cuadro 4.15. Comparación con igual trimestre del año anterior.

Observando el cuadro 4.15 se pudo pronosticar que es de esperar que el número de turistas uruguayos residentes en el exterior que ingresará a Uruguay en los próximos tres trimestres (2T09, 3T09 y 4T09) sea superior a la cantidad de turistas uruguayos residentes en el exterior ingresados en los últimos tres trimestres (2T08, 3T08 y 4T08). Cabe destacar que los trimestres tres y cuatro tendrían un aumento mínimo de turistas (0,45% y 0,19% respectivamente). En cuanto al primer trimestre de 2010 se espera que ingresen menos turistas uruguayos residentes en el exterior que los que ingresaron en el primer trimestre de 2009 (-1,69%).

4.5. Análisis de la tipología de turistas

4.5.1. Datos

La matriz de datos con la que se trabajó proviene de la Encuesta de Turismo Receptivo 2008 realizada por los encuestadores del Ministerio de Turismo y Deporte del Uruguay. Los microdatos fueron cedidos por la División Investigación y Estadísticas del Ministerio de Turismo y Deporte del Uruguay.

La ficha técnica de la Encuesta de Turismo Receptivo 2008 se encuentra en el apéndice F.1.

4.5.2. Objetivos

Para la segunda parte del trabajo, se planteó como objetivo obtener tipologías de los turistas que ingresaron a Uruguay durante el año 2008. Para ello se contó con una matriz de datos proveniente de la Encuesta de Turismo Receptivo 2008 realizada por el Ministerio de Turismo y Deporte del Uruguay. Esta matriz de datos contiene 11008 entrevistas realizadas en los 18 puntos de salida de Uruguay.



Figura 4.22. Mapa de Uruguay indicando los puntos de salida.

Para lograr el objetivo planteado se aplicaron las técnicas de análisis de componentes principales a los datos y luego se realizó análisis de clusters sobre los factores.

De acuerdo a la naturaleza de los datos, se optó por hacer dos análisis, el primero que contiene a los 11008 grupos de viaje o individuos y el segundo que contiene a 55 individuos que corresponden a los puntos de salida a lo largo de los cuatro trimestres del 2008.

Variables

Las variables se agruparon de acuerdo al tipo de información contenida de la siguiente manera:

I. **Variables de Conteo.** j_i donde j es la variable e i es el grupo encuestado.

Son las variables cuantitativas. En este grupo resultaron de interés para nuestro análisis las siguientes variables:

1. Gente: Total de integrantes del grupo de viaje.
2. Comparte: Muestra el producto de días de estadía por integrantes de cada grupo de viaje. En la generalidad de los casos esto será una multiplicación simple, pero en el caso de que no todos los integrantes del grupo hayan tenido una misma estadía, esto se refleja en esta variable.
3. Visito: Cantidad de veces que visitó Uruguay, contando esta visita.
4. Gtotal: Importe en dólares americanos del gasto total del grupo.
5. Recportap: Porcentaje de gasto con tarjeta de crédito.
6. Estadia: Duración en días del viaje a Uruguay.
7. Coef: Expansor. Se calculó con datos de la Encuesta y el Marco de Referencia, entregado por la Dirección Nacional de Migraciones. Cada trimestre tiene sus coeficientes.
8. Coeftot: Expansor. Se calculó con el producto de la variable Coef con la variable Gente. Cada trimestre tiene sus coeficientes.

II. **Variables categóricas.** j_k donde j es la variable y k son las categorías de las variables.

En este grupo resultaron de interés para nuestro análisis las siguientes variables:

1. Destino: Destino principal del viaje.
2. Lugsalid: Lugar de salida de Uruguay.

III. **Variables agregativas.** I_{kj} donde j es la variable y k son las categorías de esa variable.

$$\sum_k I_{kj} = I_j.$$

En este grupo resultaron de interés para nuestro análisis las siguientes variables:

1. Menos15: Cantidad de hombres menores de 15 años en cada grupo de viaje.
2. Entre152: Cantidad de hombres entre 15 y 29 años en cada grupo de viaje.
3. Entre306: Cantidad de hombres entre 30 y 64 años en cada grupo de viaje.
4. Mas65: Cantidad de hombres mayores de 65 años en cada grupo de viaje.
5. Fmenos15: Cantidad de mujeres menores de 15 años en cada grupo de viaje.
6. F15a29: Cantidad de mujeres entre 15 y 29 años en cada grupo de viaje.
7. F30a65: Cantidad de mujeres entre 30 y 64 años en cada grupo de viaje.

8. Fmas64: Cantidad de mujeres mayores de 65 años en cada grupo de viaje.

La suma de estas ocho variables da como resultado la variable gente.

9. Galoj: Cuanto del gasto total corresponde a alojamiento.

10. Galim: Cuanto del gasto total corresponde a alimentación.

11. Gtrans: Cuanto del gasto total corresponde a transporte.

12. Gcomp: Cuanto del gasto total corresponde a compras.

13. Gresto: Cuanto del gasto total corresponde a resto.

La suma de estas cinco variables da como resultado la variable gasto total.

Tratamiento de la información

Los microdatos se componen de 11008 encuestas realizadas a grupos de viaje en los 18 puntos de salida del país durante los cuatro trimestres del 2008, relevando 63 variables.

En esta base de datos se incluyeron los turistas uruguayos residentes en el exterior y los visitantes que permanecieron menos de 24 horas o estuvieron en tránsito a otro país. No se incluyeron los visitantes de cruceros ni quienes no hacen trámite migratorio (en el caso de la frontera seca de nuestro territorio, por ejemplo Rivera - Livramento, Chuy - Chui, Río Branco - Yaguarón)

Como el análisis se centró en caracterizar el perfil de los turistas que ingresaron a Uruguay, no se tomaron en cuenta aquellos grupos de viaje encuestados cuyo destino, motivo o localidad visitada sea "tránsito". Se realizó un estudio previo de las características de los mismos y se constató que los turistas en tránsito en el año 2008 fueron 79720 turistas constituyendo 4% del total de turistas ingresados al Uruguay (593 grupos de viaje). Se verificó que los mismos mantienen la misma composición por edad y sexo que el total de los turistas y que el gasto producido por los turistas en tránsito representó 0,6% del gasto total de los turistas en Uruguay durante el 2008, siendo esta la razón principal que llevó a no considerarlos en este análisis. Se decidió quitar los 593 grupos de viaje en tránsito por no alterar la composición de los turistas ingresados a Uruguay durante el año 2008.

Por otro lado, se debe agregar que de los 593 grupos en tránsito un 85% es de nacionalidad argentina de los cuales un 59,5% deja el país por Rivera.

Estos pasajeros ingresan al país por Paysandú o Salto y salen por Rivera en su camino a Brasil, luego en su regreso a Argentina se verifica el camino inverso.

De esta manera, los 11008 grupos de viaje encuestados se redujeron a 10415 grupos que representan un total de 1.918.160 turistas ingresados al Uruguay durante el año 2008.

Desde un principio fue necesario expandir los datos para poder trabajar con el número total de turistas ingresados a Uruguay en lugar de los grupos que fueron realmente encuestados. Para expandir las variables relacionadas con el gasto se utilizó la variable **Coef**, dado que el diseño muestral es estratificado por los 18 puntos de salida, que es donde se realizaron las encuestas. La variable **Coef** es la que indica los pesos que tienen los grupos de viaje. La variable **Coef_{tot}** es igual a $\text{Coef} \times \text{Gente}$.

4.5.3. Análisis de los datos sin agregar

4.5.3.1. Análisis de Componentes Principales

En primer lugar se aplicó análisis de componentes principales a los 10415 grupos de viaje encuestados. Las filas no tenían pesos constantes, ya que la muestra de los 10415 grupos de viaje fue estratificada por punto de salida debido a que la encuesta se realizó a un integrante del grupo de viaje al salir del país. Por lo tanto se utilizó como peso de las filas la variable expansora: **Coef**

Debido a que las 10415 filas no tenían pesos constantes, se utilizó la librería ADE4 del programa R. Dicha librería contiene la solución para este tipo de ACP en el cual las filas o columnas no tienen pesos constantes. La librería ADE4 brinda la posibilidad de elegir el peso que se le quiere dar tanto a las filas como a las columnas y se puede elegir que dicho peso sea uniforme o no. En este estudio se eligió como peso de las filas los valores de la variable expansora **Coef**. Los pesos de las columnas fueron elegidos uniformes.

Para este análisis se utilizaron las variables de conteo. Resultaron de interés para nuestro análisis las siguientes variables:

Gente: Total de integrantes del grupo de viaje.

Comparte: Muestra el producto de días de estadía por integrantes de cada grupo de viaje. En la generalidad de los casos esto será una multiplicación simple, pero en el

caso de que no todos los integrantes del grupo hayan tenido una misma estadía, esto se refleja en esta variable.

Visito: Cantidad de veces que visitó Uruguay, contando esta visita.

Gtotal: Importe en dólares americanos del gasto total del grupo.

Recportap: Porcentaje de gasto con tarjeta de crédito.

Estadia: Duración en días del viaje a Uruguay.

Coef: Expansor. Se calculó con datos de la Encuesta y el Marco de Referencia, entregado por la Dirección Nacional de Migraciones. Cada trimestre tiene sus coeficientes.

En el presente análisis se procuró determinar si es posible representar adecuadamente el conjunto de información original en un subespacio de dimensión menor. Los componentes estarán constituidos por variables construidas como combinación lineal de las originales con la menor distorsión posible.

Las variables cuantitativas consideradas tienen distintas unidades de medida, por lo que en primer lugar se estandarizaron los datos para lograr una unidad homogénea de las mismas.

A continuación se presenta el resumen de la variable Trimestre.

Trimestre	Totales
1T08	4185
2T08	2166
3T08	1983
4T08	2081

Cuadro 4.16. Número de grupos relevados por trimestre.

Luego se presenta el resumen del resto de las variables cuantitativas utilizadas.

Medidas/VARIABLES	Gente	Comparte	Visito	Gtotal	Recportap	Estadia	Coef
Mín.:	1	1	1	0	0	1	17,43
1er. Cu.:	2	6	4	378	0	4	59,21
Mediana:	2	14	6	715	0	6	69,62
Media:	2,53	20,84	4,75	1356	0,15	7,59	74,32
3er. Cu.:	3	24	6	1590	0,25	10	89,27
Máx.:	10	264	6	23652	1	61	176,63

Cuadro 4.17. Resumen de las variables.

Luego se observó la matriz de correlación que aparece en el apéndice F.2. Se calculó la matriz de correlación de las seis variables con los valores expandidos según el expansor Coef. A cada valor de cada una de las seis variables se las multiplicó por la variable Coef para obtener los valores expandidos. Estas son las variables con el sufijo Coef: Gen.Coef. Com.Coef. Vis.Coef. Gtot.Coef. Rec.Coef y Est.Coef.

También se observó que la variable Visito tiene mucha asimetría lo que puede estar indicando que la variable haya sido mal relevada. Se comprobó que la variable Visito (cantidad de veces que el grupo de viaje visitó Uruguay, contando esta visita) fue relevada de la siguiente manera:

- 1 - El grupo de viaje visitó una vez Uruguay.
- 2 - El grupo de viaje visitó dos veces Uruguay.
- 3 - El grupo de viaje visitó tres veces Uruguay.
- 4 - El grupo de viaje visitó cuatro veces Uruguay.
- 5 - El grupo de viaje visitó cinco veces Uruguay.
- 6 - El grupo de viaje visitó seis o más de seis veces Uruguay.

Entonces, se trató a la variable Visito como variable cualitativa y en principio se la quitó del análisis de componentes principales. La variable Visito fue reagrupada en tres categorías y se la volvió a incluir al final del análisis. Estas son las categorías de la nueva variable, Visito1.

1. Poco frecuente: 1 y 2 - El grupo de viaje visitó una o dos veces Uruguay.
2. Medianamente frecuente: 3 y 4 - El grupo de viaje visitó tres o cuatro veces Uruguay.
3. Muy frecuente: 5 y 6 - El grupo de viaje visitó cinco, seis o más veces Uruguay.

Luego se calculó el porcentaje de inercia acumulada y se optó por trabajar con la información contenida en tres componentes, ya que se acumuló un porcentaje alto de la inercia: 91%.

$$\lambda_1 = 0,56$$

$$\lambda_2 = 0,74$$

$$\lambda_3 = 0,91$$

$$\lambda_4 = 0,98$$

$$\lambda_5 = 1$$

Cuadro 4.18. Valores de inercia acumulada.

Luego se observó la matriz de saturación que se presenta en el apéndice y se interpretaron los tres componentes¹⁶.

El componente 1 fue formado por aquellos individuos que toman coordenadas negativas en las variables comparte, gtotal y estadia.

El primer componente indicó que las variables comparte, gtotal y estadia están relacionadas; a mayor cantidad de días de estadía, mayor cantidad de dólares gastados.

El componente 2 fue formado por aquellos individuos que toman coordenadas altas y negativas en la variable recportap.

El componente 3 fue formado por aquellos individuos que toman coordenadas altas y positivas en la variable gente.

A continuación se presenta la calidad de representación de las variables en el plano principal.

¹⁶ Ver Apéndice F.3.

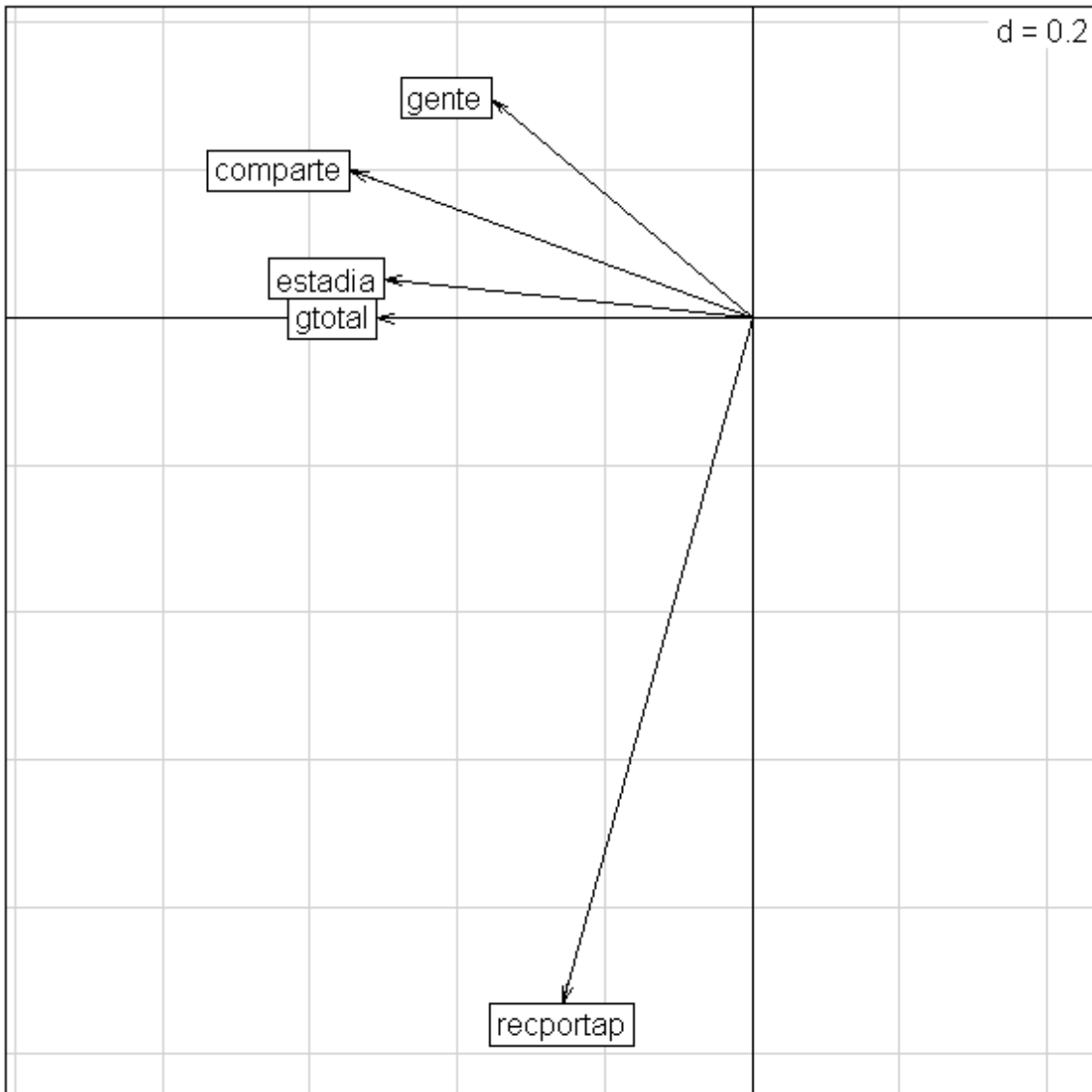


Figura 4.23. Calidad de representación de las variables en el plano factorial. (F_1 y F_2)

La calidad de representación de los individuos en 3 componentes es correcta ya que solo 133 individuos o grupos de viaje quedaron mal representados, constituyendo un 1,28% de los 10415 grupos de viaje.

Por lo tanto, se resolvió trabajar con 3 componentes.

4.5.3.2. Análisis de Clusters

Nuestro análisis se basó en encontrar tipologías para los 10415 grupos de viaje que fueron encuestados al salir de Uruguay durante 2008, en base a los factores, agrupando dichos grupos de viaje en clusters o conglomerados. Para el cálculo de las distancias entre observaciones se usaron las coordenadas factoriales.

Como insumo del análisis de clusters se tomó la matriz de coordenadas de los 10415 individuos obtenida en los 3 primeros componentes del ACP.

Debido al alto número de observaciones (10415) se utilizaron los métodos de Análisis de Clusters No Jerárquicos. Para ello se utilizó la función k-medias que se usa para asignar individuos a un número fijo de clusters cuyas características no se conocen aún pero que se basan en un conjunto de variables especificadas.

En la función “cascadeKM” participan la función k-medias y el índice Calinski. La misma se utilizó para determinar el número de clusters.

Se probó con k = 3, 4, 5 y 6 y el valor máximo del índice Calinski fue 5963,5 que corresponde a tres clusters o 5928,9 que corresponde a cinco clusters. Por lo tanto, se pudo haber trabajado con tres o cinco clusters.

	3 clusters	4 clusters	5 clusters	6 clusters
Calinski	5963,5	5809,3	5928,9	5811,9

Cuadro 4.19. Salida del índice Calinski.

	3 clusters	5 clusters	Nro. Turistas	% Turistas	Cant. USD gastados	% USD gastado
1	1058	1285	284.407	14,8%	299.595.747	29,2%
2	7233	330	101.443	5,3%	177.251.848	17,3%
3	2124	4908	634.440	33,1%	240.826.864	23,5%
4		1599	292.420	15,2%	153.924.651	15,0%
5		2293	605.427	31,6%	154.203.516	15,0%

Cuadro 4.20. Tamaño y composición de los clusters.

Se decidió trabajar con 5 clusters ya que proporciona una estructura de grupos más pareja. Si se trabajara con 3 clusters uno de los clusters tendría casi el 70% de los individuos.

El cluster 1 contiene 1285 grupos de viaje que representan 284.407 (14,8%) turistas que ingresaron a Uruguay durante el año 2008 que gastaron USD 299.595.747 (29,2%).

El cluster 2 contiene 330 grupos de viaje que representan 101.443 (5,3%) turistas que ingresaron a Uruguay durante el año 2008 que gastaron USD 177.251.848 (17,3%).

El cluster 3 contiene 4908 grupos de viaje que representan 634.440 (33,1%) turistas que ingresaron a Uruguay durante el año 2008 que gastaron USD 240.826.864 (23,5%).

El cluster 4 contiene 1599 grupos de viaje que representan 292.420 (15,2%) turistas que ingresaron a Uruguay durante el año 2008 que gastaron USD 153.924.651 (15%).

El cluster 5 contiene 2293 grupos de viaje que representan 605.427 (31,6%) turistas que ingresaron a Uruguay durante el año 2008 que gastaron USD 154.203.516 (15%).

Mediante los diagramas de caja se caracterizaron los cinco clusters.

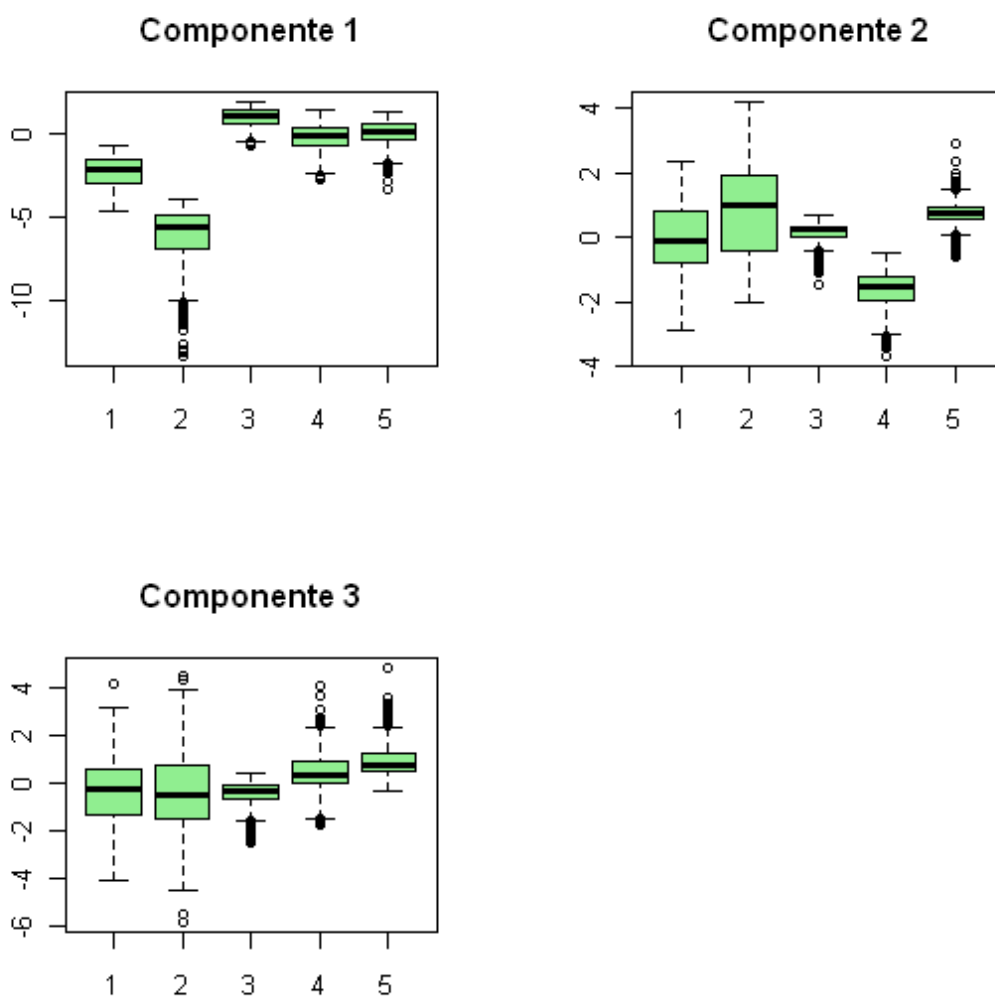


Figura 4.24. Diagrama de cajas por cluster para los tres componentes principales.

Cluster 1

Son aquellos grupos de viaje que presentaron gasto medio y duración de estadía intermedia. También tuvieron niveles medios de porcentaje del gasto total con tarjeta de crédito y de integrantes del grupo de viaje. Además se observó que hay una gran dispersión de este cluster en relación al segundo y tercer componente.

Cluster 2

Aquí se encuentran los grupos de viaje que presentaron los mayores niveles de gasto total y días de estadía. Además tuvieron los niveles más bajos de porcentaje del gasto total con tarjeta de crédito. Presentaron niveles bajos de integrantes del grupo de viaje. También se observó que hay una gran dispersión de este cluster en relación al segundo y tercer componente. Este cluster es el que tiene menor frecuencia, tiene 330 grupos de viaje.

Cluster 3

Fue formado por los grupos de viaje que presentaron los menores niveles de gasto total y días de estadía. También tuvieron niveles medios de porcentaje del gasto total con tarjeta de crédito y de integrantes del grupo de viaje. Este cluster es el que tiene mayor frecuencia con 4908 grupos de viaje.

Cluster 4

Este cluster reúne a los grupos de viaje que presentaron niveles intermedios de gasto total y días de estadía. También presentaron los niveles más altos de porcentaje del gasto total con tarjeta de crédito y niveles medios de integrantes del grupo de viaje.

Cluster 5

Son aquellos grupos de viaje que presentaron gasto medio y duración de estadía intermedia. También tuvieron niveles medios de porcentaje del gasto total con tarjeta de crédito y niveles altos de integrantes del grupo de viaje.

En resumen, el **cluster 1** reúne a los grupos de viaje con niveles medios de gasto total, días de estadía, porcentaje del gasto total con tarjeta de crédito y de integrantes del grupo de viaje. Es un cluster intermedio.

El **cluster 2** reúne a los grupos de viaje con niveles altos de gasto total y días de estadía y con niveles bajos de porcentaje del gasto total con tarjeta de crédito y de integrantes del grupo de viaje. Este es el cluster con menor frecuencia, tiene 330 grupos de viaje.

El **cluster 3** reúne a los grupos de viaje con niveles bajos de gasto total y días de estadía y con niveles medios de porcentaje del gasto total con tarjeta de crédito y de integrantes del grupo de viaje. Este es el cluster con mayor frecuencia, tiene 4908 grupos de viaje.

El **cluster 4** reúne a los grupos de viaje con niveles medios de gasto total, días de estadía y de integrantes del grupo de viaje y niveles altos de porcentaje del gasto total con tarjeta de crédito.

El **cluster 5** reúne a los grupos de viaje con niveles medios de gasto total, días de estadía y de porcentaje del gasto total con tarjeta de crédito y niveles altos de integrantes del grupo de viaje.

La variable Visito fue quitada del análisis al principio y a continuación se analiza la variable Visito1.

Cluster/Visito1	1	2	3	Total
1	234	299	752	1285
2	38	40	252	330
3	706	1061	3141	4908
4	269	357	973	1599
5	248	438	1607	2293
Total	1495	2195	6725	10415

Cuadro 4.21. Tabla de contingencia por modalidad de la variable Visito1 según los 5 clusters.

Cluster/Visito1	1	2	3	Total
1	18,2%	23,3%	58,5%	100%
2	11,5%	12,1%	76,4%	100%
3	14,4%	21,6%	64,0%	100%
4	16,8%	22,3%	60,9%	100%
5	10,8%	19,1%	70,1%	100%

Cuadro 4.22. Porcentajes fila de la tabla de contingencia por modalidad de la variable Visito1 según los 5 clusters.

Se observó que en los cinco clusters hay mayor porcentaje de gente que visitó Uruguay con mucha frecuencia (5, 6 o más veces).

Luego le sigue la categoría medianamente frecuente (3 o 4 veces) y finalmente la categoría poco frecuente (1 o 2 veces).

Cluster/Visito1	1	2	3
1	15,7%	13,6%	11,2%
2	2,5%	1,8%	3,7%
3	47,2%	48,3%	46,7%
4	18,0%	16,3%	14,5%
5	16,6%	20,0%	23,9%
Total	100%	100%	100%

Cuadro 4.23. Porcentajes columna de la tabla de contingencia por modalidad de la variable Visito1 según los 5 clusters.

Se observó que la categoría poco frecuente tiene mayor presencia en el cluster 3 y menor presencia en el cluster 2. La categoría medianamente frecuente tiene mayor presencia en el cluster 3 y menor presencia en el cluster 2. Finalmente, la categoría muy frecuente también tiene mayor presencia en el cluster 3 y menor presencia en el cluster 2. Esto se debe a que el cluster 3 es el que tiene mayor número de grupos de viaje (4908) mientras que el cluster 2 tiene la menor cantidad de grupos de viaje (330).

4.5.4. Análisis de los datos agregados

4.5.4.1. Análisis de Componentes Principales

En esta parte se decidió hacer un análisis desde el punto de vista de la evolución temporal de los puntos de salida en los que son encuestados los grupos de viaje a lo largo de los cuatro trimestres del año 2008. Para eso fue necesario crear objetos comparables en los cuatro trimestres ya que los grupos de viaje encuestados son únicos y todos distintos entre sí, es decir un mismo grupo de viaje no es encuestado en los cuatro trimestres. Cada grupo de viaje fue encuestado una única vez al salir del país. Por lo tanto, se decidió comparar los 18 puntos de salida a lo largo de los cuatro trimestres del 2008.

De estos 18 puntos de salida solo 14 tuvieron actividad para el primer, segundo y cuarto trimestre. Los cuatro puntos de salida que no tuvieron actividad son: Fray Bentos, Otros, Aeropuerto Colonia y Piriápolis. En el tercer trimestre Artigas tampoco

tuvo actividad. Por lo tanto, se contó con 55 puntos de salida para los cuatro trimestres del 2008.

Para esto fue necesario expandir los datos, utilizando las variables **Coef** y **Coefot**. Luego se agregaron los grupos de viaje para obtener totales en función del punto de salida y del trimestre en que salieron de Uruguay, logrando así que las 55 filas tengan peso constante. Para esto se utilizó la librería SURVEY del programa R que permitió agregar todos los grupos de viaje pertenecientes a cada punto de salida, usando las variables mencionadas anteriormente para expandir los datos. Esto se repitió para cada punto de salida y luego para cada trimestre, obteniendo así un valor para cada variable en los 55 individuos. Se repitió este mismo procedimiento para cada una de las 14 variables que se utilizaron. Luego se calcularon ratios sobre esta matriz de datos de dimensión 55 x 14 y se realizó ACP común.

Las variables que se incluyeron por considerarse de interés en este análisis fueron las variables agregativas de edad-sexo y gasto, la variable categórica destino y la variable estadía.

Las variables agregativas son:

1. Menos15: Cantidad de hombres menores de 15 años en cada grupo de viaje.
2. Entre152: Cantidad de hombres entre 15 y 29 años en cada grupo de viaje.
3. Entre306: Cantidad de hombres entre 30 y 64 años en cada grupo de viaje.
4. Mas65: Cantidad de hombres mayores de 65 años en cada grupo de viaje.
5. Fmenos15: Cantidad de mujeres menores de 15 años en cada grupo de viaje.
6. F15a29: Cantidad de mujeres entre 15 y 29 años en cada grupo de viaje.
7. F30a65: Cantidad de mujeres entre 30 y 64 años en cada grupo de viaje.
8. Fmas64: Cantidad de mujeres mayores de 65 años en cada grupo de viaje.

La suma de estas ocho variables da como resultado la variable gente.

9. Galoj: Cuanto del gasto total corresponde a alojamiento.
10. Galim: Cuanto del gasto total corresponde a alimentación.
11. Gtrans: Cuanto del gasto total corresponde a transporte.
12. Gcomp: Cuanto del gasto total corresponde a compras.
13. Gresto: Cuanto del gasto total corresponde a resto.

La suma de estas cinco variables da como resultado la variable gasto total.

Lugar de salida de Uruguay

Valor	Categoría	Valor	Categoría
1	Colonia	10	Salto
2	Puerto	11	Aeropuerto Colonia
3	Carrasco	12	Piriápolis
4	Chuy	13	Rivera
5	Carmelo	14	Nueva Palmira
6	Punta del Este	15	Artigas
7	Fray Bentos	16	Bella Unión
8	Paysandú	17	Melo - Aceguá
9	Otros	18	Río Branco

Cuadro 4.24. Puntos de salida del grupo de viaje encuestado.

Nota: De estos 18 puntos de salida solo 14 tuvieron actividad para el primer, segundo y cuarto trimestre. Los cuatro puntos de salida que no tuvieron actividad son: Fray Bentos, Otros, Aeropuerto Colonia y Piriápolis. En el tercer trimestre Artigas tampoco tuvo actividad. Por lo tanto, se contó con 55 puntos de salida para los cuatro trimestres del 2008. Son 14 puntos de salida por 4 trimestres = 56 puntos de salida, menos Artigas 3T. En total se contó con 55 puntos de salida.

Estadia: Duración en días del viaje a Uruguay.

Destino principal del viaje

Valor	Categoría
1	Punta del Este
2	Colonia
3	Montevideo
4	Costa de Oro
5	Piriápolis
6	Costas Oceánicas
7	Litoral Termal
8	Tránsito
9	Otros

Cuadro 4.25. Destino principal elegido por el grupo de viaje encuestado.

Luego de que se prepararon los datos en cuatro tablas - una por cada trimestre, se intentó aplicarle Análisis de Componentes Principales (ACP) para caracterizar cada trimestre pero se observó que eran muy pocas filas (14 puntos de salida) para la

cantidad de variables (14). El ACP es una técnica que no permite trabajar con mayor número de columnas que filas.

Para resolver este inconveniente, se unieron las cuatro tablas en una, de dimensión 55x14, que contiene los 55 puntos de salida en los cuatro trimestres y las 14 variables. Con el objetivo de analizar la evolución temporal de los puntos de salida en los que son encuestados los grupos de viaje se aplicó ACP para obtener factores y luego se aplicó Análisis de Cluster. De esta manera los 55 puntos de salida quedaron divididos en distintos clusters con características similares entre los puntos de salida dentro de cada cluster y características diferentes entre los puntos de salida de distintos clusters.

Cada uno de los 55 individuos representa un punto de salida de Uruguay en los cuales se encuestaron a los grupos de viaje. El mismo punto de salida se repite en cada uno de los cuatro trimestres del año 2008 como se presenta a continuación.

1	1T1	Colonia	15	1T2	Colonia	29	1T3	Colonia	42	1T4	Colonia
2	2T1	Puerto	16	2T2	Puerto	30	2T3	Puerto	43	2T4	Puerto
3	3T1	Carrasco	17	3T2	Carrasco	31	3T3	Carrasco	44	3T4	Carrasco
4	4T1	Chuy	18	4T2	Chuy	32	4T3	Chuy	45	4T4	Chuy
5	5T1	Carmelo	19	5T2	Carmelo	33	5T3	Carmelo	46	5T4	Carmelo
6	6T1	Punta del Este	20	6T2	Punta del Este	34	6T3	Punta del Este	47	6T4	Punta del Este
7	8T1	Paysandú	21	8T2	Paysandú	35	8T3	Paysandú	48	8T4	Paysandú
8	10T1	Salto	22	10T2	Salto	36	10T3	Salto	49	10T4	Salto
9	13T1	Rivera	23	13T2	Rivera	37	13T3	Rivera	50	13T4	Rivera
10	14T1	Nueva Palmira	24	14T2	Nueva Palmira	38	14T3	Nueva Palmira	51	14T4	Nueva Palmira
11	15T1	Artigas	25	15T2	Artigas	X	X	XXX	52	15T4	Artigas
12	16T1	Bella Unión	26	16T2	Bella Unión	39	16T3	Bella Unión	53	16T4	Bella Unión
13	17T1	Melo – Aceguá	27	17T2	Melo – Aceguá	40	17T3	Melo – Aceguá	54	17T4	Melo – Aceguá
14	18T1	Río Branco	28	18T2	Río Branco	41	18T3	Río Branco	55	18T4	Río Branco

Cuadro 4.26. Enumeración del punto de salida por trimestre.

Posteriormente, con el objetivo de reducir la cantidad de variables debido a la baja cantidad de individuos se redujeron las variables de la siguiente manera:

Variables de Edad

Se colapsaron las categorías de sexo para obtener cuatro variables de edad:

Men15.rat = Menos15 + Fmenos15 = Porcentaje de personas menores de 15 años en cada grupo de viaje.

$X_{1529.rat} = Entre_{152} + F_{15a29}$ = Porcentaje de personas entre 15 y 29 años en cada grupo de viaje.

$X_{3064.rat} = Entre_{306} + F_{30a65}$ = Porcentaje de personas entre 30 y 64 años en cada grupo de viaje.

$Mas_{65.rat} = Mas_{65} + F_{mas64}$ = Porcentaje de personas mayores de 65 años en cada grupo de viaje.

Variables de Gasto

Se colapsaron las cinco categorías en las tres más representativas de la composición del gasto de los turistas ingresados a Uruguay en el 2008 de esta manera:

$Galoj.rat = Galoj$ = Porcentaje del gasto total que corresponde a alojamiento.

$Galim.rat = Galim$ = Porcentaje del gasto total que corresponde a alimentación.

$Gotr.rat = Gtrans + Gcomp + Gresto$ = Porcentaje del gasto total que corresponde a transporte, compras y resto.

Variable Destino

Se colapsaron las categorías de esta variable en seis categorías:

Destino1	Punta del Este
Destino2	Colonia
Destino3	Montevideo
Destino4	Costa de Oro, Piriápolis y Otros
Destino6	Costas Oceánicas
Destino7	Litoral Termal

Cuadro 4.27. Reclasificación de la variable destino.

Se calcularon ratios a partir de estas nuevas 13 variables, menos la variable estadía que se expresó en días promedio de estadía por grupo por punto de salida, logrando de esta manera que las 55 filas tengan peso constante. Esto se debe a que se agregaron todos los grupos de viaje encuestados por punto de salida utilizando como peso el expansor Coef, consiguiendo que los 55 puntos de salida tengan peso constante. Luego se procedió a aplicarle ACP a la matriz de 55 individuos x 14 variables.

1	Destino1 - Punta del Este
2	Destino2 – Colonia
3	Destino3 – Montevideo
4	Destino4 - Costa de Oro, Piriápolis y Otros
5	Destino6 - Costas Oceánicas
6	Destino7 - Litoral Termal
7	Men15.rat - Porcentaje de personas menores de 15 años en cada grupo de viaje.
8	X1529.rat - Porcentaje de personas entre 15 y 29 años en cada grupo de viaje.
9	X3064.rat - Porcentaje de personas entre 30 y 64 años en cada grupo de viaje.
10	Mas.65rat - Porcentaje de personas mayores de 65 años en cada grupo de viaje.
11	Galoj.rat - Porcentaje del gasto total que corresponde a alojamiento.
12	Galim.rat - Porcentaje del gasto total que corresponde a alimentación.
13	Gotr.rat - Porcentaje del gasto total que corresponde a transporte, compras y resto.
14	Estadia.prom.pg - Duración promedio en días del viaje a Uruguay, por grupo de viaje, por punto de salida.

Cuadro 4.28. Tabla de 14 variables utilizadas en el ACP.

Luego se observó la matriz de correlación y se concluyó que no existen altas correlaciones entre las variables¹⁷.

Se observó el porcentaje de inercia acumulada y se resolvió trabajar con 4 ó 5 componentes, acumulando el 67% o el 75% de la inercia total.

Componente 1	0,22
Componente 2	0,42
Componente 3	0,57
Componente 4	0,67
Componente 5	0,75
Componente 6	0,83
Componente 7	0,89
Componente 8	0,93
Componente 9	0,96
Componente 10	0,98
Componente 11	1
Componente 12	1
Componente 13	1
Componente 14	1

Cuadro 4.29. Tabla de inercia acumulada.

Se observó que el individuo 52 tiene norma alta (93,31) con respecto a las normas del resto de los individuos. El individuo 52 corresponde al punto de salida Artigas en el cuarto trimestre. Esto indica que este punto de salida se encuentra mucho más lejos del baricentro que el resto de los individuos y que el mismo puede ser considerado un outlier. Entonces, se corrió un nuevo ACP sin el individuo 52 y se observó que los

¹⁷ Ver Apéndice F.4.

valores de la norma de los 54 individuos son semejantes, es decir ningún valor sobresale.

El punto de salida 52 (Artigas 4T) es particular debido a los siguientes motivos:

- Los grupos de viaje que salieron por Artigas en el cuarto trimestre se quedaron en promedio la menor cantidad de días de estadía por grupo: 2,3.
- Por este punto de salida no salieron personas menores de 15 años, ni mayores de 65 años.
- Los grupos de viaje que salieron por Artigas en el cuarto trimestre eligieron solo dos destinos: Costa de Oro, Piriápolis y Otros y Litoral Termal.
- Los grupos de viaje que salieron por Artigas en el cuarto trimestre solo gastaron en alimentación y en otros. Gastaron muy poco en alojamiento, a diferencia de los grupos de viaje que salieron por Artigas en el primer y segundo trimestre.

Luego se observó la matriz de saturación¹⁸ que se presenta en el apéndice y se interpretaron los cinco componentes.

El componente 1 fue formado por aquellos individuos que tomaron coordenadas positivas en la variable destino7 (Litoral Termal) y por aquellos individuos que tomaron coordenadas negativas en la variable gotr.rat (otros gastos).

El componente 2 fue formado por aquellos individuos que tomaron coordenadas positivas en las variables estadia, X1529.rat (turistas entre 15 y 29 años), destino1 (Punta del Este), destino6 (Costas Oceánicas) y galoj.rat (gasto en alojamiento) y por aquellos individuos que tomaron coordenadas negativas en las variables destino4 (Costa de Oro, Pirlápolis y Otros) y galim.rat (gasto en alimentación).

El segundo componente diferencia los destinos Punta de Este y Costas Oceánicas del destino Costa de Oro, Pirlápolis y Otros. También diferencia aquellos individuos que gastaron mayoritariamente en alojamiento de aquellos que gastaron mayoritariamente en alimentación.

El componente 3 fue formado por aquellos individuos que tomaron coordenadas positivas en la variable X3064.rat (turistas entre 30 y 64 años) y por aquellos

¹⁸ Ver Apéndice F.5.

individuos que tomaron coordenadas negativas en la variable mas.65rat (turistas mayores de 65 años).

El tercer componente diferencia los turistas con edad media de los turistas con edad más avanzada.

El componente 4 fue formado por aquellos individuos que tomaron coordenadas positivas en la variable men15.rat (turistas con menos de 15 años de edad) y por aquellos individuos que tomaron coordenadas negativas en la variable destino2 (Colonia).

El componente 5 fue formado por aquellos individuos que tomaron coordenadas negativas en la variable destino3 (Montevideo).

A continuación se presenta la calidad de representación de las variables en el plano principal.

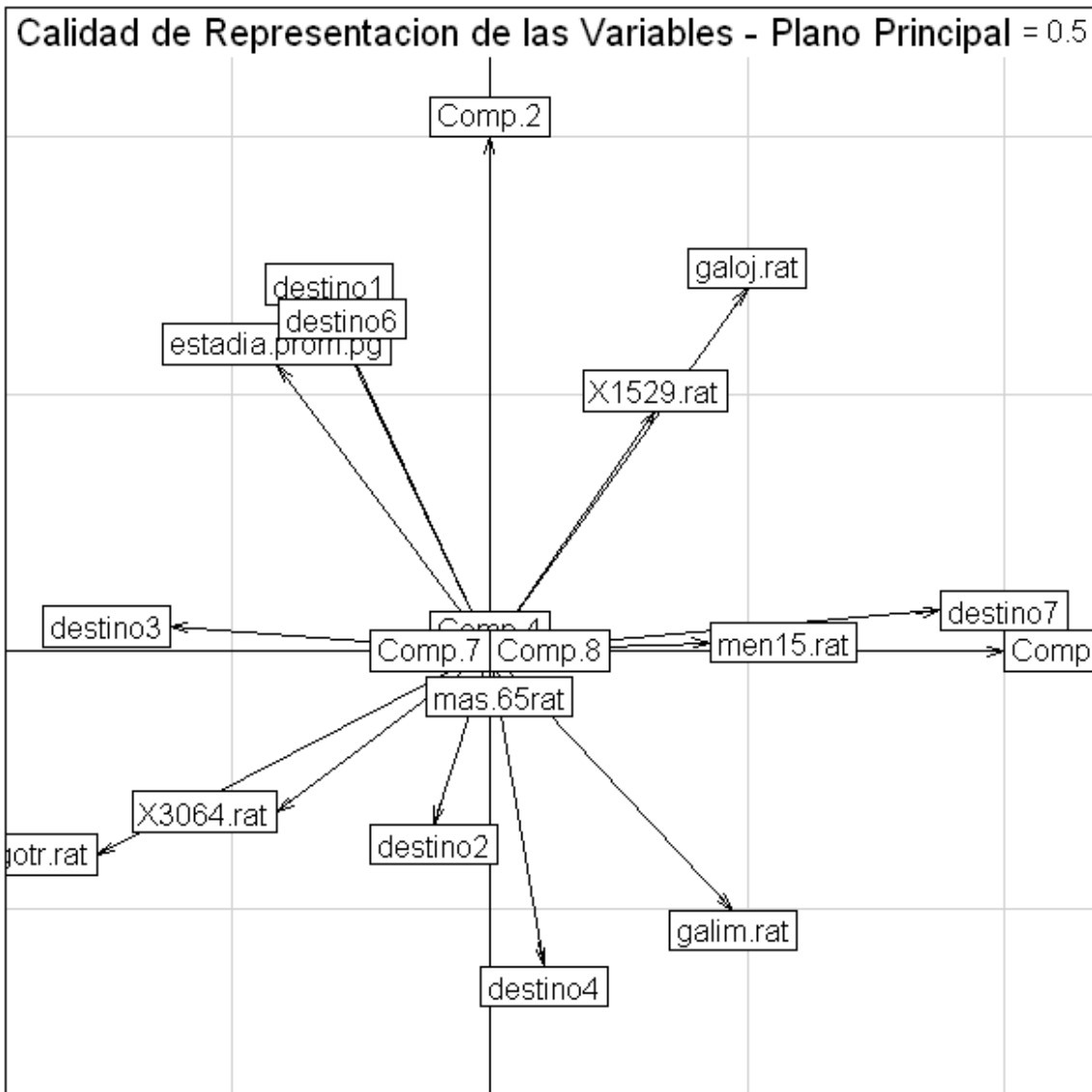


Figura 4.25. Calidad de representación de las variables en el plano principal.

La calidad de representación de los individuos en 5 componentes es correcta ya que siete individuos quedaron mal representados, de los cuales cinco tuvieron coseno2 cercano a 0,5. De esta manera, solo dos individuos quedaron mal representados, el 27 y el 29 que corresponden a los puntos de salida Melo-Aceguá2T y Colonia3T.

Por lo tanto, se resolvió trabajar con 5 componentes.

A continuación se presenta la proyección de los individuos en el plano principal.

Proyección de Individuos - Plano Principal

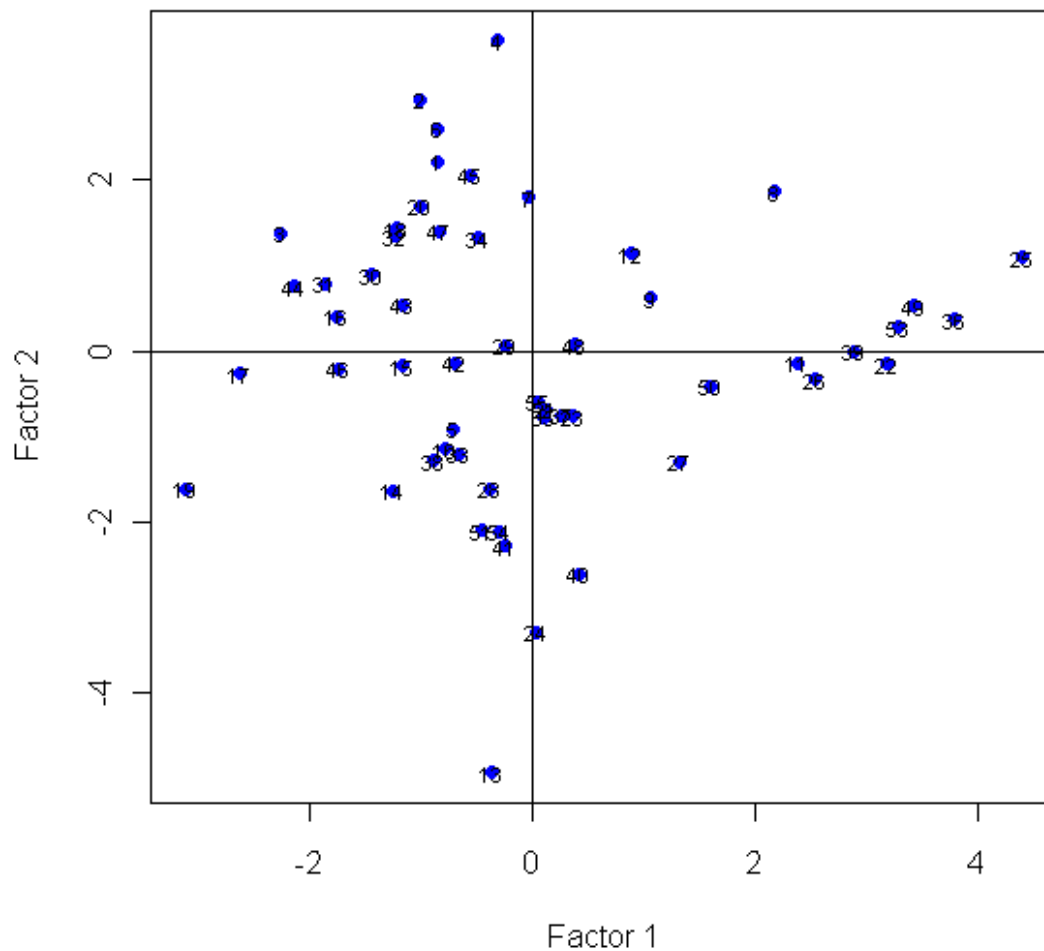


Figura 4.26. Proyección de los individuos en el plano principal.

4.5.4.2. Análisis de Clusters

En los respectivos algoritmos de unión que se desarrollaron se utilizó la distancia euclidiana entre individuos, con la cual se construyeron las Matrices de Disimilitud.

Existen varios algoritmos de unión, tales como los Algoritmos de Ward, Vecino más cercano y Vecino más lejano entre otros.

El método de agregación por el que se optó es el Algoritmo de Ward debido a que este método está basado en la varianza y en su descomposición y su uso se entendió mas adecuado. En resumen, el objetivo de este método es el de minimizar la varianza (o dispersión) intragrupos y maximizar la misma entre grupos.

El uso de indicadores ayudó en la selección óptima de la cantidad de grupos a retener. Para ello se contó con tres indicadores denominados R^2 , Pseudo F y Pseudo t^2 .

Para decidir la cantidad de clusters con la cual trabajar, se basó tanto en los indicadores como en el dendrograma. De acuerdo a los indicadores anteriormente descritos se tomó la decisión de la cantidad de clusters a utilizar, para luego realizar una caracterización de los individuos dentro de cada uno de los clusters elegidos.

Como insumo del análisis de clusters se tomó la matriz de coordenadas de los 54 individuos obtenida en los 5 primeros componentes del ACP.

La historia de aglomeración según el algoritmo de Ward se presenta en el apéndice¹⁹. Se pudo haber elegido 3 o 5 clusters con este algoritmo. A continuación se presenta el dendrograma.

¹⁹ Ver Apéndice F.6.

Algoritmo de Ward

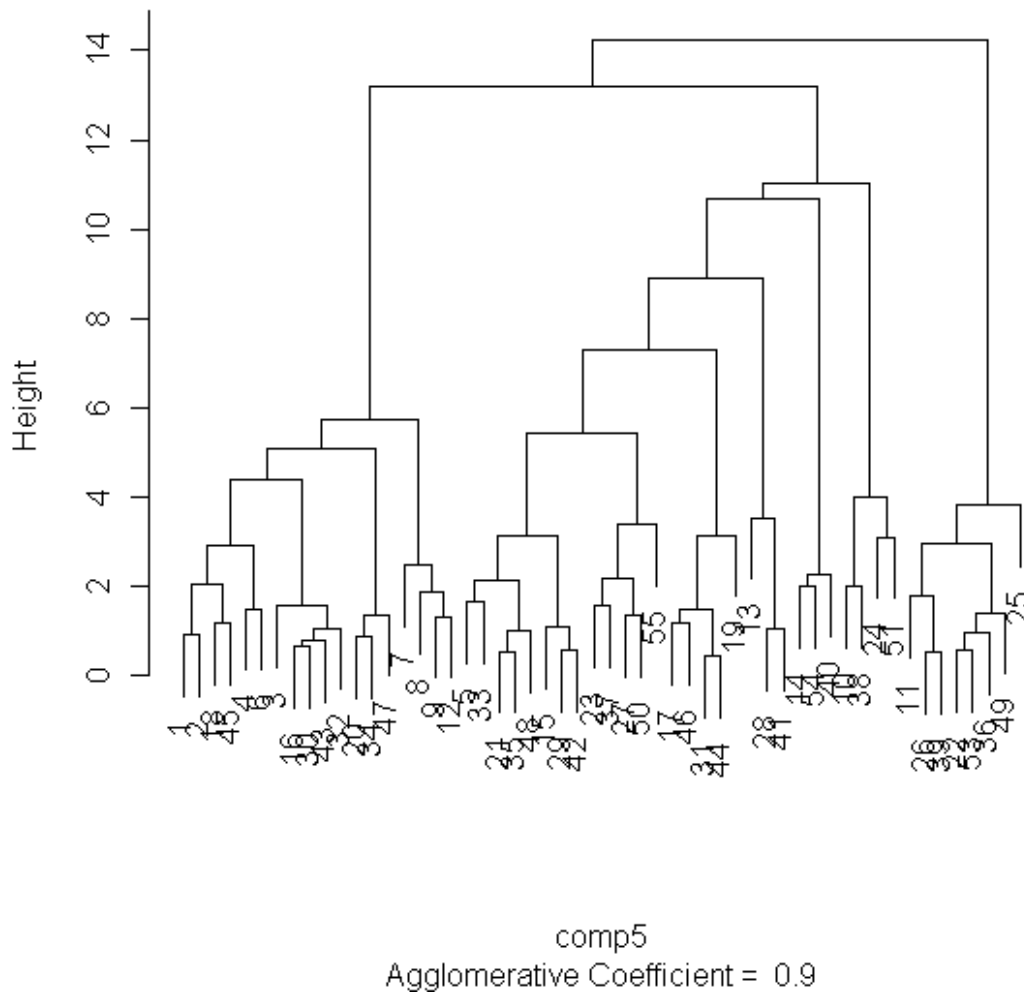


Figura 4.27. Dendrograma según algoritmo de Ward.

Se decidió elegir 3 clusters ya que proporciona una estructura de grupos más pareja.

El tamaño de los 3 clusters es el siguiente.

	3 clusters	Nro. Turistas	% Turistas	Cant. USD gastados	% USD gastado
1	18	878.296	45,8%	534.062.374	52,1%
2	28	921.993	48,1%	465.832.850	45,4%
3	8	117.220	6,1%	25.653.848	2,5%

Cuadro 4.30. Tamaño y composición de los clusters.

El grupo 1 contiene 18 puntos de salida que representan 878.296 (45,8%) turistas que ingresaron a Uruguay durante el año 2008 que gastaron USD 534.062.374 (52,1%).

El grupo 2 contiene 28 puntos de salida que representan 921.993 (48,1%) turistas que ingresaron a Uruguay durante el año 2008 que gastaron USD 465.832.850 (45,4%).

El grupo 3 contiene 8 puntos de salida que representan 117.220 (6,1%) turistas que ingresaron a Uruguay durante el año 2008 que gastaron USD 25.653.848 (2,5%).

Mediante los diagramas de caja se caracterizaron los tres clusters.

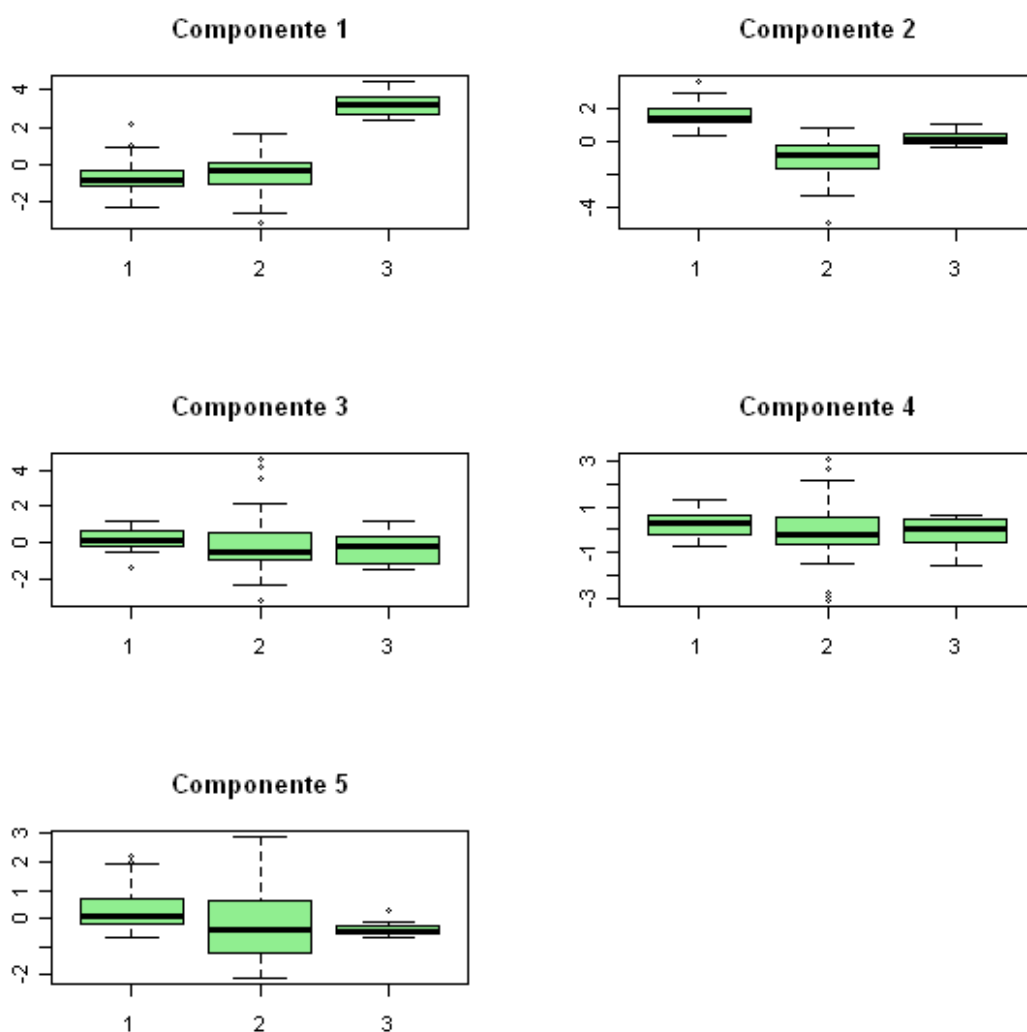


Figura 4.28. Diagrama de cajas por cluster para los cinco componentes principales, según el algoritmo de Ward.

Cluster 1

Son aquellos individuos que presentaron el mayor gasto en otros rubros. También son aquellos individuos con mayor cantidad de turistas que visitaron Punta del Este y Costas Oceánicas y que en promedio se quedaron la mayor cantidad de días de estadía y que gastaron mayoritariamente en alojamiento. Presentaron la mayor cantidad de turistas en edad adulta (30 a 64 años) y la menor cantidad de turistas en edad avanzada (más de 65 años). También son aquellos individuos que presentaron la mayor cantidad de turistas menores de 15 años y la menor cantidad de turistas que visitaron Colonia y Montevideo.

Cluster 2

Son aquellos individuos que presentaron gasto alto en otros rubros. También son aquellos individuos con menor cantidad de turistas que visitaron Punta del Este y Costas Oceánicas y que en promedio se quedaron la menor cantidad de días de estadía, con el menor gasto en alojamiento. Presentaron cantidad intermedia de turistas en edad adulta (30 a 64 años) y cantidad intermedia de turistas en edad avanzada (más de 65 años).

También presentaron la menor cantidad de turistas menores de 15 años y la mayor cantidad de turistas que visitaron Colonia y Montevideo.

Además se observó que hay dispersión de este cluster en relación a los cinco componentes. Este cluster es el que tiene mayor frecuencia, tiene 28 puntos de salida.

Cluster 3

Son aquellos individuos con mayor cantidad de turistas que visitaron el Litoral Termal y que presentaron bajo gasto en otros rubros. También son aquellos individuos con cantidad intermedia de turistas que visitaron Punta del Este y Costas Oceánicas y con gasto intermedio en alojamiento. Presentaron la menor cantidad de turistas en edad adulta (30 a 64 años) y la mayor cantidad de turistas en edad avanzada (más de 65 años). También presentaron cantidad intermedia de turistas menores de 15 años y cantidad intermedia de turistas que visitaron Colonia y Montevideo. Este cluster es el que tiene menor frecuencia, tiene 8 puntos de salida.

En resumen, el **cluster 1** agrupa a los individuos que contienen a los turistas que más visitaron Punta del Este y Costas Oceánicas y que más gasto realizaron. También contienen a la mayoría de los turistas en edad adulta y a la minoría de los turistas que visitaron Colonia y Montevideo y a la minoría de los adultos mayores.

El **cluster 2** agrupa a los individuos que contienen a los turistas que más visitaron Montevideo y Colonia y realizaron un gasto intermedio, con nivel intermedio de todas las edades de los turistas. También contienen a la minoría de los turistas que visitaron Punta del Este y Costas Oceánicas.

El **cluster 3** agrupa a los individuos que contienen a los turistas que más visitaron el Litoral Termal. También contienen a la mayoría de los turistas en edad avanzada y a la minoría del resto de las edades de los turistas. Tiene un nivel intermedio de turistas que visitaron Punta del Este, Costas Oceánicas, Montevideo y Colonia. Son los turistas que realizaron un menor nivel de gasto.

A continuación se presentan los nombres de los puntos de salida que pertenecen a cada cluster.

Cluster 1

1	Colonia 1T	16	Puerto 2T
2	Puerto 1T	18	Chuy 2T
3	Carrasco 1T	20	Punta del Este 2T
4	Chuy 1T	30	Puerto 3T
6	Punta del Este 1T	32	Chuy 3T
7	Paysandú 1T	34	Punta del Este 3T
8	Salto 1T	43	Puerto 4T
9	Rivera 1T	45	Chuy 4T
12	Bella Unión 1T	47	Punta del Este 4T

La mayoría de los turistas que visitaron Punta del Este y Costas Oceánicas salieron por los puntos de salida contenidos en el cluster 1. Estos turistas se encuentran en la franja etaria 30 a 64 años y su nivel de gasto fue mayor que el del resto de los turistas.

Cluster 2

5	Carmelo 1T	33	Carmelo 3T
10	Nueva Palmira 1T	35	Paysandú 3T
13	Melo - Aceguá 1T	37	Rivera 3T
14	Río Branco 1T	38	Nueva Palmira 3T
15	Colonia 2T	40	Melo - Aceguá 3T
17	Carrasco 2T	41	Río Branco 3T
19	Carmelo 2T	42	Colonia 4T
21	Paysandú 2T	44	Carrasco 4T
23	Rivera 2T	46	Carmelo 4T
24	Nueva Palmira 2T	48	Paysandú 4T
27	Melo - Aceguá 2T	50	Rivera 4T
28	Río Branco 2T	51	Nueva Palmira 4T
29	Colonia 3T	54	Melo - Aceguá 4T
31	Carrasco 3T	55	Río Branco 4T

La mayoría de los turistas que visitaron Montevideo y Colonia salieron por los puntos de salida contenidos en el cluster 2. Su nivel de gasto fue medio.

Cluster 3

11	Artigas 1T
22	Salto 2T
25	Artigas 2T
26	Bella Unión 2T
36	Salto 3T
39	Bella Unión 3T
49	Salto 4T
53	Bella Unión 4T

La mayoría de los turistas que visitaron el Litoral Termal salieron por los puntos de salida contenidos en el cluster 3. Estos turistas se encuentran en la franja etaria de más de 65 años y su nivel de gasto fue menor que el del resto de los turistas.

4.5.4.3. Evolución Temporal de los Puntos de Salida por Trimestre

Para caracterizar a los clusters a lo largo de los cuatro trimestres del 2008 se presentan las siguientes tablas.

Cluster/Trimestre	1T08	2T08	3T08	4T08	Total
1	9	3	3	3	18
2	4	8	8	8	28
3	1	3	2	2	8
Total	14	14	13	13	54

Cuadro 4.31. Tabla de los puntos de salida por trimestre, según tipología de Método de Ward.

Cluster/Trimestre	1T08	2T08	3T08	4T08	Total
1	50,0%	16,7%	16,7%	16,7%	100%
2	14,3%	28,6%	28,6%	28,6%	100%
3	12,5%	37,5%	25,0%	25,0%	100%

Cuadro 4.32. Porcentajes fila de los puntos de salida por trimestre, según tipología de Método de Ward.

Cluster/Trimestre	1T08	2T08	3T08	4T08
1	64,3%	21,4%	23,1%	23,1%
2	28,6%	57,1%	61,5%	61,5%
3	7,1%	21,4%	15,4%	15,4%
Total	100%	100%	100%	100%

Cuadro 4.33. Porcentajes columna de los puntos de salida por trimestre, según tipología de Método de Ward.

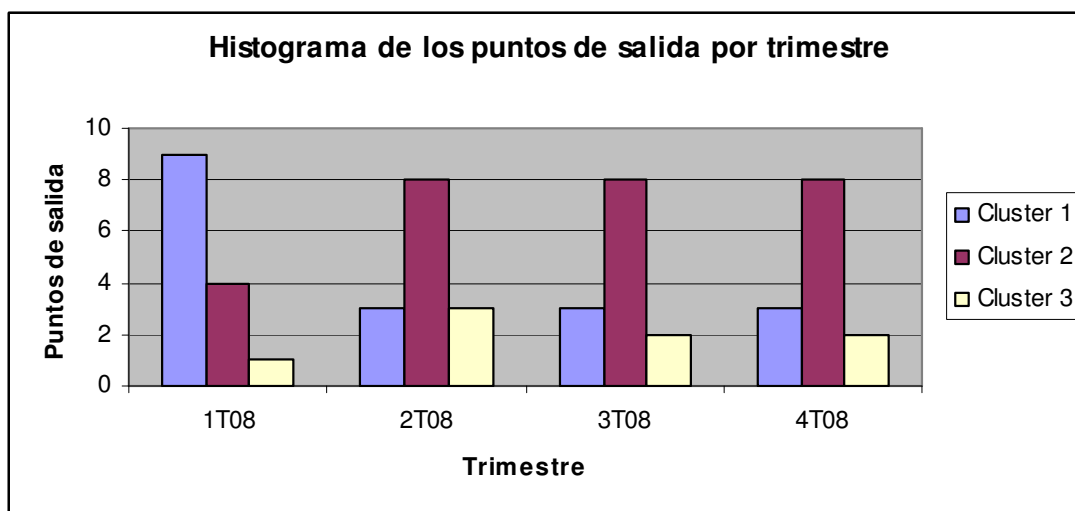


Figura 4.29. Histograma de los puntos de salida por trimestre

Se observó que el tercer y cuarto trimestre presentaron los mismos puntos de salida en cada uno de los tres clusters, indicando un mismo perfil del turista que salió de Uruguay en el tercer y cuarto trimestre.

Se observó que a lo largo del año 2008 para el cluster 1, 50% de los puntos de salida se presentaron en el primer trimestre. Estos son aquellos puntos por los que salieron los turistas que visitaron Punta del Este y Costas Oceánicas durante el primer trimestre. Los tres trimestres siguientes presentaron la misma cantidad de puntos de salida, es decir un 16,7%.

En el segundo cluster prevalecieron los puntos de salida correspondientes al segundo, tercer y cuarto trimestre del 2008. Estos son aquellos puntos por los que salieron los turistas que visitaron Montevideo y Colonia.

Además se observó en el tercer cluster que contiene a aquellos puntos de salida por los que salieron los turistas que visitaron el Litoral Termal que son menos numerosos en el primer trimestre (verano) y más numerosos en el segundo trimestre (otoño).

Luego se hizo el análisis de la evolución de la cantidad de turistas según los tres clusters.

Cluster/Trimestre	1T08	2T08	3T08	4T08	Total
1	659.277	50.537	64.376	104.108	878.298
2	30.448	221.746	270.124	399.679	921.997
3	2.966	35.470	36.457	42.324	117.217
Total	692.691	307.753	370.957	546.111	1.917.512

Cuadro 4.34. Total expandido de turistas correspondiente a los puntos de salida por trimestre, según tipología de Método de Ward.

Cluster/Trimestre	1T08	2T08	3T08	4T08	Total
1	75,1%	5,8%	7,3%	11,9%	100%
2	3,3%	24,1%	29,3%	43,3%	100%
3	2,5%	30,3%	31,1%	36,1%	100%

Cuadro 4.35. Porcentajes fila del total expandido de turistas correspondiente a los puntos de salida por trimestre, según tipología de Método de Ward.

Se observó que aquellos grupos de viaje que salieron por los puntos de salida del cluster 1 (que en su mayoría visitaron Punta del Este y Costas Oceánicas) reúnen 75,1% de los turistas en el primer trimestre.

Los grupos de viaje del cluster 2 (que en su mayoría visitaron Montevideo y Colonia) reúnen 43,3% de los turistas en el cuarto trimestre y 24,1% y 29,3% en el segundo y tercer trimestre respectivamente.

Los grupos de viaje del cluster 3 (que en su mayoría visitaron el Litoral Termal) reúnen alrededor del 30% de los turistas en cada uno de los trimestres 2, 3 y 4.

Cluster/Trimestre	1T08	2T08	3T08	4T08
1	95,2%	16,4%	17,4%	19,1%
2	4,4%	72,1%	72,8%	73,2%
3	0,4%	11,5%	9,8%	7,8%
Total	100%	100%	100%	100%

Cuadro 4.36. Porcentajes columna del total expandido de turistas correspondiente a los puntos de salida por trimestre, según tipología de Método de Ward.

También se observó que en el primer trimestre el 95,2% de los turistas corresponden a los grupos de viaje del cluster 1 (que en su mayoría visitaron Punta del Este y Costas Oceánicas).

En el segundo, tercer y cuarto trimestre, alrededor del 70% de los turistas de cada trimestre corresponden a los grupos de viaje del cluster 2 (que en su mayoría visitaron Montevideo y Colonia).

Finalmente se hizo el análisis de la evolución del gasto total según los tres clusters.

Cluster/Trimestre	1T08	2T08	3T08	4T08	Total
1	424.692.919	21.626.981	32.697.960	55.040.261	534.058.120
2	8.587.215	83.776.561	163.429.528	210.034.754	465.828.058
3	572.544	7.206.509	7.989.069	9.886.322	25.654.445
Total	433.852.678	112.610.050	204.116.557	274.961.337	1.025.540.623

Cuadro 4.37. Gasto total expandido correspondiente a los puntos de salida por trimestre, según tipología de Método de Ward.

Cluster/Trimestre	1T08	2T08	3T08	4T08	Total
1	79,5%	4,0%	6,1%	10,3%	100%
2	1,8%	18,0%	35,1%	45,1%	100%
3	2,2%	28,1%	31,1%	38,5%	100%

Cuadro 4.38. Porcentajes fila del gasto total expandido correspondiente a los puntos de salida por trimestre, según tipología de Método de Ward.

Se observó que 79,5% del gasto total realizado por los turistas que en su mayoría visitaron Punta del Este y Costas Oceánicas se efectuó en el primer trimestre.

Los turistas que en su mayoría visitaron Montevideo y Colonia gastaron 45,1% en el cuarto trimestre, 35,1% en el tercer trimestre y 18,0% en el segundo trimestre.

Los turistas que en su mayoría visitaron el Litoral Termal gastaron alrededor del 30% en cada uno de los trimestres 2, 3 y 4.

Si se compara el cluster 2 a lo largo de los cuatro trimestres del 2008 su comportamiento difiere en términos de turistas y de gasto realizado. Se observó una tendencia creciente de turistas y de gasto realizado en el cluster 2 del primer trimestre hacia el cuarto trimestre del 2008.

Cluster/Trimestre	1T08	2T08	3T08	4T08
1	97,9%	19,2%	16,0%	20,0%
2	2,0%	74,4%	80,1%	76,4%
3	0,1%	6,4%	3,9%	3,6%
Total	100%	100%	100%	100%

Cuadro 4.39. Porcentajes columna del gasto total expandido correspondiente a los puntos de salida por trimestre, según tipología de Método de Ward.

En el primer trimestre, 97,9% del gasto total se realizó por los turistas que en su mayoría visitaron Punta del Este y Costas Oceánicas.

Los turistas que en su mayoría visitaron Montevideo y Colonia gastaron alrededor de 75%-80% en el segundo, tercer y cuarto trimestre.

Como sugerencia se plantea otro enfoque para el análisis de los datos agregados: Análisis Factorial Múltiple o incluso Análisis Multiway. El análisis multiway es una rama especializada dentro de la estadística multivariada que amplía los métodos frecuentemente empleados para datos a dos vías como ser análisis factorial, análisis de componentes principales y análisis de correspondencias. En este caso se estudiarían los datos del tipo individuos (grupos de viaje) x variables x tiempo (trimestres).

Para el Análisis Factorial Múltiple se pueden hacer tres ACPs:

1. 55 individuos (puntos de salida) x 8 variables de edad-sexo.
2. 55 individuos (puntos de salida) x 5 variables de gasto.
3. 55 individuos (puntos de salida) x 6 variables de destino.

Y luego extraer un componente de cada ACP y juntarlos ponderando cada componente por el inverso del primer valor propio de cada ACP para que sean comparables. De esta manera se obtiene un análisis factorial con 3 factores que salen cada uno de su respectivo ACP y son comparables.

A esta nueva tabla de datos se le puede aplicar Análisis de Clusters para la elaboración de una tipología que se podría comparar con la que se obtuvo en este estudio.

CAPÍTULO 5: CONCLUSIONES Y RECOMENDACIONES

Como se mencionó al principio, el turismo es uno de los factores más importantes en la generación de producto de una economía nacional.

Para este trabajo se plantearon dos objetivos. El primero, estimar el número de turistas que ingresarán al país y el segundo, segmentarlos en distintos perfiles.

Para el primer objetivo se consideraron tres series temporales de ingreso de turistas a Uruguay. Para el segundo objetivo se contó con los datos de la Encuesta de Turismo Receptivo 2008 del Ministerio de Turismo y Deporte del Uruguay.

Las series de ingresos con las que se trabajó son las siguientes:

- 1) Cantidad de turistas que ingresaron a Uruguay trimestralmente, en el período 1993.I-2009.I.
- 2) Cantidad de turistas que ingresaron a Uruguay mensualmente, en el período 1996.01-2009.07.
- 3) Cantidad de turistas uruguayos residentes en el exterior que ingresaron a Uruguay trimestralmente, en el período 1996.I-2009.I.

Para lograr el primer objetivo planteado se estimaron tres modelos SARIMA-IA (modelo estacional autorregresivo y de medias móviles con análisis de intervención):

1) Modelo 1

$$\log(\hat{y}_t) = 0,53\log(y_{t-1}) + \log(y_{t-4}) - 0,53\log(y_{t-5}) + 0,08Turismo - 0,43Tc2002$$

El modelo ARIMA $(1,0,0) (0,1,0)_{(4)}$ fue intervenido por:

- **Turismo:** Es una variable indicatriz que toma valor uno en los trimestres en que cae la semana de turismo y cero en el resto de los trimestres.
- **Tc2002:** Es el cambio transitorio de la crisis económica regional que comenzó en Diciembre 2001 y que afectó a la baja el ingreso de turistas a Uruguay a partir del primer trimestre del año 2002 en adelante. Como su efecto es transitorio, el mismo va disminuyendo a medida que transcurre el tiempo hasta desaparecer.

2) Modelo 2

$$\log(\hat{y}_t) = 0,84\log(y_{t-1}) + \log(y_{t-2}) - 0,84\log(y_{t-3}) + 0,59\varepsilon_{t-2} + 0,26\text{Turismo} - 0,31\text{Tc_Ene02} + 0,30\text{Tc_May02}$$

El modelo ARIMA $(1,0,0) (0,1,1)_{(12)}$ fue intervenido por:

- **Turismo:** Es una variable indicatriz que toma valor uno en los meses en que cae la semana de turismo y cero en el resto de los meses.
- **Tc_Ene02:** Es el cambio transitorio de la crisis económica regional que comenzó en Diciembre 2001 y que afectó a la baja el ingreso de turistas a Uruguay a partir de enero de 2002 en adelante. Como su efecto es transitorio, el mismo va disminuyendo a medida que transcurre el tiempo hasta desaparecer.
- **Tc_May02:** Es el cambio transitorio del retiro de capitales depositados en la banca uruguaya por parte de los argentinos que afectó positivamente sobre el número de turistas que ingresaron al país provenientes de Argentina a partir de mayo de 2002 en adelante. Como su efecto es transitorio, el mismo va disminuyendo a medida que transcurre el tiempo hasta desaparecer.

3) Modelo 3

$$\log(\hat{y}_t) = \log(y_{t-4}) - 0,32\varepsilon_{t-1} + 0,77\varepsilon_{t-4} - 0,25\varepsilon_{t-5} - 0,33\text{Ls2002}$$

El modelo ARIMA $(0,0,1) (0,1,1)_{(4)}$ fue intervenido por:

- **Ls2002:** Es el cambio de nivel de la crisis económica regional que comenzó en Diciembre 2001 y que afectó a la baja el ingreso de turistas uruguayos residentes en el exterior a Uruguay a partir del primer trimestre del año 2002 en adelante. Como su efecto es permanente el mismo se mantiene hasta el final de la serie.

Se advirtió que la crisis del 2002 tuvo efecto transitorio para los modelos 1 y 2 mientras que para el modelo 3 tuvo un efecto permanente. Otra discrepancia que se encontró es que la variable indicatriz Turismo incidió en el ingreso de turistas a Uruguay pero no afectó particularmente al ingreso de turistas uruguayos residentes en el exterior a Uruguay.

El estudio permitió concluir que si se considera el modelo 2 las predicciones indicaron que el número de turistas que ingresará a Uruguay en los próximos doce meses disminuirá (6,42%), según cuadro 4.10, mientras que si se considera el modelo 1 las predicciones indicaron que el número de turistas que ingresará a Uruguay en los próximos cuatro trimestres aumentará (6,24%), según cuadro 4.4.

Se recuerda que la serie mensual no incluye a los turistas uruguayos residentes en el exterior mientras que la serie trimestral si los incluye.

Si se considera el modelo 3 es de esperar que el número de turistas uruguayos residentes en el exterior que ingresará a Uruguay en los próximos cuatro trimestres aumente ligeramente (0,1%), según cuadro 4.15.

A continuación se presentan los valores reales de turistas que ingresaron a Uruguay en el segundo, tercer y cuarto trimestre del 2009 y en el primer trimestre del 2010 provistos por el Ministerio de Turismo y Deporte del Uruguay.

Trimestre	Predicción	Valor real	Int. Inf. 95	Int. Sup. 95	Error	Error en %
2T-2009	358.618	379.392	299.762	429.030	-20.774	-5,5%
3T-2009	388.732	354.170	317.182	476.421	34.562	9,8%
4T-2009	573.284	638.777	464.770	707.135	-65.493	-10,3%
1T-2010	794.501	755.279	642.947	981.779	39.222	5,2%
Total	2.115.135	2.127.618			-12.483	-0,6%

Cuadro 4.40. Comparación de los valores reales con las predicciones del modelo 1.

Se observó que en 2T09 y 4T09 el modelo 1 subestimó el número de turistas mientras que en 3T09 y 1T10 el modelo 1 sobreestimó el número de turistas. El error porcentual global para los cuatro trimestres fue -0,6%.

Trimestre	2008/2009	2009/2010	Tasa de Crecimiento Real
2T	317.096	379.392	19,6%
3T	380.560	354.170	-6,9%
4T	566.788	638.777	12,7%
1T	726.441	755.279	4,0%

Cuadro 4.41. Tasa de Crecimiento Real

La predicción fuera de la muestra del modelo 1 dio como resultado que era de esperar que el número de turistas que ingresaría a Uruguay en los próximos cuatro trimestres fuera superior a la cantidad de turistas ingresados en los últimos cuatro trimestres (2T08, 3T08, 4T08 y 1T09). Contando con los valores reales de ingreso de turistas

para 2T09, 3T09, 4T09 y 1T10, salvo en 3T09, el número de turistas ingresados a Uruguay fue mayor en los tres trimestres restantes.

Para el segundo objetivo se contó con los datos de la Encuesta de Turismo Receptivo 2008 del Ministerio de Turismo y Deporte del Uruguay. Se plantearon dos análisis, el primero que consistió en aplicar ACP y Análisis de Clusters a la matriz de datos que contiene 10415 observaciones y 5 variables: Gente, Comparte, Gtotal, Recportap y Estadia con el objetivo de encontrar tipologías de los turistas que ingresaron a Uruguay durante el año 2008. El segundo análisis consistió en aplicar ACP y Análisis de Clusters a la matriz de datos que contiene 55 individuos y 14 variables: 4 variables de edad, 3 variables de gasto, días de estadía promedio por grupo y 6 variables de destino con el objetivo de analizar la evolución temporal de los 55 individuos a lo largo de los cuatro trimestres del 2008.

Los resultados encontrados permitieron concluir que en el análisis de los datos sin agregar se encontraron cinco grupos con la siguiente composición.

El cluster 1 reúne a los grupos de viaje con niveles medios de gasto total, días de estadía, porcentaje del gasto total con tarjeta de crédito y de integrantes del grupo de viaje. Es un cluster intermedio. La frecuencia de este cluster es 1285 grupos de viaje.

El cluster 2 reúne a los grupos de viaje con niveles altos de gasto total y días de estadía y con niveles bajos de porcentaje del gasto total con tarjeta de crédito y de integrantes del grupo de viaje. Este es el cluster con menor frecuencia, tiene 330 grupos de viaje.

El cluster 3 reúne a los grupos de viaje con niveles bajos de gasto total y días de estadía y con niveles medios de porcentaje del gasto total con tarjeta de crédito y de integrantes del grupo de viaje. Este es el cluster con mayor frecuencia, tiene 4908 grupos de viaje.

El cluster 4 reúne a los grupos de viaje con niveles medios de gasto total, días de estadía y de integrantes del grupo de viaje y niveles altos de porcentaje del gasto total con tarjeta de crédito. La frecuencia de este cluster es 1599 grupos de viaje.

El cluster 5 reúne a los grupos de viaje con niveles medios de gasto total, días de estadía y de porcentaje del gasto total con tarjeta de crédito y niveles altos de integrantes del grupo de viaje. La frecuencia de este cluster es 2293 grupos de viaje.

También se analizó la variable Visito y como recomendación para el futuro, se sugiere relevar la variable Visito en forma cuantitativa, es decir relevar el número de veces que el grupo de viaje visitó Uruguay, contando esta visita o bien relevarla en forma cualitativa, teniendo en cuenta las tres categorías propuestas: poco frecuente, medianamente frecuente y muy frecuente.

En el análisis de los datos agregados se encontraron tres grupos con la siguiente composición.

El cluster 1 agrupa a los individuos que contienen a los turistas que más visitaron Punta del Este y Costas Oceánicas y que más gasto realizaron. También contienen a la mayoría de los turistas en edad adulta y a la minoría de los turistas que visitaron Colonia y Montevideo y a la minoría de los adultos mayores. Este cluster contiene 18 puntos de salida.

El cluster 2 agrupa a los individuos que contienen a los turistas que más visitaron Montevideo y Colonia y realizaron un gasto intermedio, con nivel intermedio de todas las edades de los turistas. También contienen a la minoría de los turistas que visitaron Punta del Este y Costas Oceánicas. Este cluster contiene 28 puntos de salida.

El cluster 3 agrupa a los individuos que contienen a los turistas que más visitaron el Litoral Termal. También contienen a la mayoría de los turistas en edad avanzada y a la minoría del resto de las edades de los turistas. Tiene un nivel intermedio de turistas que visitaron Punta del Este, Costas Oceánicas, Montevideo y Colonia. Son los turistas que realizaron un menor nivel de gasto. Este cluster contiene 8 puntos de salida.

La mayoría de los turistas que visitaron Punta del Este y Costas Oceánicas salieron por los puntos de salida contenidos en el cluster 1. Estos turistas se encuentran en la franja etaria 30 a 64 años y su nivel de gasto fue mayor que el del resto de los turistas. La mayoría de los turistas que visitaron Montevideo y Colonia salieron por los puntos de salida contenidos en el cluster 2. Su nivel de gasto fue medio.

La mayoría de los turistas que visitaron el Litoral Termal salieron por los puntos de salida contenidos en el cluster 3. Estos turistas se encuentran en la franja etaria de más de 65 años y su nivel de gasto fue menor que el del resto de los turistas.

Finalmente se analizó la evolución temporal de los 55 individuos ó puntos de salida a lo largo de los cuatro trimestres del 2008. Se observó que el tercer y cuarto trimestre presentaron los mismos puntos de salida en cada uno de los tres grupos, indicando un mismo perfil del turista que salió de Uruguay en el tercer y cuarto trimestre del 2008.

Se observó que a lo largo del año 2008 para el cluster 1, 50% de los puntos de salida corresponden al primer trimestre. Estos son aquellos puntos de salida por los que salieron los turistas que visitaron Punta del Este y Costas Oceánicas durante el primer trimestre de 2008.

En el segundo cluster prevalecieron los puntos de salida correspondientes al segundo, tercer y cuarto trimestre del 2008. Estos son aquellos puntos de salida por los que salieron los turistas que visitaron Montevideo y Colonia.

Además se observó que el tercer cluster contiene a aquellos puntos de salida por los que salieron los turistas que visitaron el Litoral Termal durante el 2008 que son menos numerosos en el primer trimestre (verano) y más numerosos en el segundo trimestre (otoño).

De acuerdo a lo mencionado se observó una clara estacionalidad de los distintos destinos turísticos: Punta del Este y Costas Oceánicas en el primer trimestre y Litoral Termal en el segundo trimestre. Esto es algo que todos los operadores turísticos deberían tener en cuenta a la hora de promover sus productos turísticos locales.

En base a lo expuesto se apreció que cada análisis dio una visión distinta de la información analizada. Mientras que el análisis de los datos sin agregar brindó un enfoque clásico en el cual se obtuvieron 5 clusters para caracterizar distintos perfiles de los turistas que ingresaron en el 2008, el análisis de los datos agregados ofreció una perspectiva de la evolución temporal de los turistas que visitaron Uruguay durante el año 2008. Esto también permitió observar la estacionalidad estival para aquellos turistas que vinieron a veranear a Uruguay en el correr del primer trimestre del 2008, que mayoritariamente fueron turistas que salieron de Uruguay por los puntos contenidos en el cluster 1 del análisis de datos agregados.

Para poder efectuar cada análisis, previamente se utilizaron dos herramientas distintas para tratar el tema de los pesos no constantes de los grupos de viaje. Se buscó darle a cada grupo de viaje la misma importancia dentro de cada uno de los dos análisis.

Los grupos de viaje o filas no tenían pesos constantes, ya que la muestra de los 10415 grupos de viaje fue estratificada por punto de salida, por lo tanto se utilizó como peso de las filas la variable expansora: Coef.

En el análisis de los datos sin agregar se utilizó la librería ADE4 del programa R. Dicha librería contiene la solución para este tipo de ACP en el cual las filas o columnas no tienen pesos constantes. La librería ADE4 brinda la posibilidad de elegir el peso que se le quiere dar tanto a las filas como a las columnas y se puede elegir que dicho peso sea uniforme o no. En este estudio se eligió como peso de las filas los valores de la variable expansora Coef y los pesos de las columnas fueron elegidos uniformes.

En el análisis de los datos agregados se agregaron los grupos de viaje para obtener totales en función del punto de salida y del trimestre en que salieron de Uruguay, logrando así que las 55 filas tengan peso constante. Para esto se utilizó la librería SURVEY del programa R que permitió agregar todos los grupos de viaje pertenecientes a un punto de salida, usando las variables expansoras Coef y Coeftot para expandir los datos. Esto se repitió para cada punto de salida y luego para cada trimestre, obteniendo así un valor para cada variable en los 55 individuos.

El análisis de los datos agregados supuso más trabajo preliminar para el procesamiento de los datos que el análisis de los datos sin agregar, aunque este último presentó la dificultad de trabajar con una cantidad mayor de grupos de viaje.

Por lo tanto, en base a los resultados cuantitativos que se obtuvieron en la segunda parte del trabajo, para los turistas ingresados a Uruguay en el 2008 se consideró enfocar las políticas en:

- Aumentar la cantidad de turistas cuyo destino sea Punta del Este y Costas Oceánicas, siendo éstos los destinos en los que más gastaron los turistas que ingresaron en el primer trimestre.
- Para aumentar el gasto, ampliar las opciones de consumo en Montevideo y Colonia, siendo éstos los destinos más visitados en el segundo, tercer y cuarto trimestre.
- Aumentar el número de turistas que ingresan al Litoral Termal, destino en el que menos gastaron los turistas.
- Aumentar los días de estadía promedio por grupo para lograr que los turistas gasten más. Una posibilidad para aumentar la estadía de los turistas sería fomentando la promoción de paquetes de x noches más una de regalo.

Nótese que este estudio debería repetirse todos los años ya que la parte de estimación del número de turistas que ingresarán a Uruguay fue realizada con datos hasta mediados del 2009 y el análisis de la tipología de los turistas que ingresaron a Uruguay fue realizado en base a los microdatos del año 2008 y por lo tanto dichas conclusiones son válidas solamente para ese año.

BIBLIOGRAFÍA

[1] Brida, J.G., Lanzilotta, B. y Risso, W. (2008), "Turismo y crecimiento económico: el caso de Uruguay". Documento presentado en las XXIII Jornadas de Economía del Banco Central del Uruguay.

[2] Espasa, A. y Canelo, J. (1993), "Métodos cuantitativos para el análisis de la coyuntura económica. Capítulo 10 – Un análisis econométrico del turismo en España: Implicaciones para el estudio sectorial de las exportaciones y algunas consideraciones de política económica". Alianza Editorial.

[3] Robano, V. (2000), "Determinantes del turismo receptivo en Uruguay". Documento presentado en las XV Jornadas de Economía del Banco Central del Uruguay.

[4] Armellini, M. e Isabella, F. (2003), "Turismo receptivo en Uruguay: una evaluación del aporte al producto, el empleo y las remuneraciones". Documento presentado en las XVIII Jornadas del Banco Central del Uruguay.

[5] Glejberman, D. y Sosa, V. (2003), Memorando: "Predicción del número de visitantes del turismo receptivo para los próximos trimestres". División Investigación y Estadísticas, Ministerio de Turismo y Deporte del Uruguay, Montevideo, Uruguay.

[6] Mantero, R., Perelmuter, N. y Sueiro, I. (2004), "Determinantes económicos del turismo receptivo en Uruguay: ¿Aporta información relevante un análisis desagregado?". CINVE.

[7] Gil-Alana, L.A., Cuñado, J. y Pérez de Gracia, F. (2008), "Tourism in the Canary Islands: Forecasting Using Several Seasonal Time Series Models". Journal of Forecasting, John Wiley & Sons, Ltd.

[8] Glejberman, D. (2004), "Tipologías del Turismo Receptivo". División Investigación y Estadísticas, Ministerio de Turismo y Deporte del Uruguay, Montevideo, Uruguay.

[9] Brida, J.G., Bukstein, D., Garrido, N., Tealde, E., Zapata, S. (2010), "Impacto económico del turismo de cruceros. Un análisis del gasto de los cruceristas que visitan el caribe colombiano", Documento de Investigación N° 58, Facultad de Administración y Ciencias Sociales, Universidad ORT, Montevideo, Uruguay.

- [10] Box, G. y Jenkins, G. (1976), "Time series analysis: forecasting and control". Editorial Holden Day.
- [11] Chatfield, C. (1989), "The Analysis of Time Series: An Introduction". Editorial Chapman & Hall.
- [12] Hamilton, J. (1994), "Time series analysis". Princeton University Press.
- [13] Jarque, C. y Bera, A. (1980), "Efficient tests for Normality, Homocedasticity and Serial Independence of Regression Residuals". Economics Letters. Number 6.
- [14] Peña, D. (2005), "Análisis de Series Temporales". Alianza Editorial.
- [15] Blanco, J.C. (2006), "Introducción al análisis multivariado". Iesta, Facultad de Ciencias Económicas y de Administración, Universidad de la República, Montevideo, Uruguay.
- [16] Universidad de Cádiz, Área de Informática, (Enero 2010), <http://www2.uca.es/serv/ai/formacion/spss/Imprimir/21conglk.pdf>
- [17] Ouellette, M., Durand, S. y Legendre, P. (Enero 2010) - Universidad de Montreal, <http://cc.oulu.fi/~jarioksa/softhelp/vegan/html/cascadeKM.html>
- [18] Calinski, T. y Harabasz, J. (1974), "A Dendrite Method for Cluster Analysis". Communications in Statistics, 3, 1-27.
- [19] Milligan y Cooper (1985), "Simulation Study".

APÉNDICE A: METODOLOGÍA

A.1. Metodología aplicada a las series de tiempo univariadas

A.1.1. Modelización SARIMA(p,d,q)(P,D,Q)

Una serie temporal o cronológica es una secuencia de datos, observaciones o valores, medidos en determinados momentos del tiempo, ordenados cronológicamente y, normalmente, espaciados entre sí de manera uniforme. El análisis de series temporales comprende métodos que ayudan a interpretar este tipo de datos, extrayendo información representativa, como también la posibilidad de extrapolar y predecir su comportamiento futuro. De hecho uno de los usos más habituales de las series de datos temporales es su análisis para predicción.

El objetivo del análisis de una serie de tiempo es el conocimiento de su patrón de comportamiento, para así poder predecir su evolución en el futuro, suponiendo que las condiciones futuras de la dinámica de la serie no variarán respecto al pasado.

Una de las técnicas estadísticas que puede emplearse en este contexto es la metodología de los modelos ARIMA formalizada por Box y Jenkins (1976) [10]. “Este enfoque parte del hecho de que la serie temporal que se trata de predecir es generada por un proceso estocástico cuya naturaleza puede ser caracterizada mediante un modelo. Para efectuar la estimación de un modelo ARIMA se requiere de una serie temporal que cuente con un elevado número de observaciones. Básicamente, la metodología Box-Jenkins consiste en encontrar un modelo matemático que represente el comportamiento de una serie temporal de datos y permita hacer predicciones únicamente introduciendo el período de tiempo correspondiente.” Chatfield (1989) [11] En los modelos ARIMA univariantes se explica el comportamiento de una serie temporal a partir de observaciones pasadas de la propia serie y a partir de los errores pasados de previsión.

La especificación del modelo univariado ARIMA estacional multiplicativo de series temporales, queda planteado según la siguiente ecuación:

$$\Phi_p(L^s)\phi_p(L)\nabla_s^D\nabla^d Y_t = \Theta_q(L^s)\theta_q(L)\varepsilon_t \quad (\text{A.1})$$

A partir de la fórmula (A.1) se logra modelizar la dependencia regular y estacional en forma conjunta. El componente regular, refiere a la dependencia entre observaciones

consecutivas y el componente estacional, corresponde a la frecuencia de las observaciones separadas por s períodos, por ejemplo: trimestres, meses, etc.

Los componentes de la ecuación (A.1) se interpretan de la siguiente manera:

p y **P**: orden del polinomio autorregresivo de la parte regular y estacional.

d y **D**: número de diferencias de orden regular y estacional, que la serie necesita para ser estacionaria.

q y **Q**: orden del polinomio de medias móviles de la parte regular y estacional.

Dónde:

$\Phi_p(L^s) = (1 - \Phi_1 L^s - \dots - \Phi_p L^{sP})$ es el operador AR estacional de orden P.

$\phi_p(L) = (1 - \phi_1 L - \phi_2 L^2 - \dots - \phi_p L^p)$ es el operador AR regular de orden p.

$\nabla_s^D \nabla^d = (1 - L^s)^D (1 - L)^d$ representan las diferencias estacional y regular respectivamente.

$\Theta_Q(L^s) = (1 - \Theta_1 L^s - \dots - \Theta_Q L^{sQ})$ es el operador media móvil estacional de orden Q.

$\theta_q(L) = (1 - \theta_1 L - \theta_2 L^2 - \dots - \theta_q L^q)$ es el operador media móvil regular de orden q.

ε_t es un proceso llamado ruido blanco o innovación, que en general se supone gaussiano.

Un proceso ruido blanco es un proceso estacionario definido por las siguientes condiciones:

1. $E(\varepsilon_t) = 0 \quad t = 1, 2, \dots$
2. $Var(\varepsilon_t) = \sigma^2 \quad t = 1, 2, \dots$
3. $Cov(\varepsilon_t, \varepsilon_{t-k}) = 0 \quad k = \pm 1, \pm 2, \dots$

A.1.2. Supuestos del modelo univariado

Estacionariedad

La metodología Box - Jenkins tiene como supuesto básico la estacionariedad de la serie analizada. Una serie es estacionaria si la media y la varianza de la misma son constantes en el tiempo.

Además, si para un intervalo de tiempo de longitud k se calcula la autocorrelación de dos variables integrantes del proceso estocástico, el resultado sólo depende del

tamaño del intervalo que las separa (k). Si esto sucede, la serie se denomina estacionaria en sentido débil.

Los supuestos que implican la estacionariedad del modelo son:

1. Media constante en el tiempo.

$$\mu_t = E(X_t)$$

2. Varianza constante en el tiempo (homocedasticidad) y finita

$$\sigma_t^2 = \text{Var}(X_t) \leq \infty$$

3. La autocovarianza y la autocorrelación entre dos variables sólo depende del intervalo que las separa (de tamaño k).

$$\text{Cov}(X_t, X_{t+k}) = \text{Cov}(X_{t+j}, X_{t+j+k}) = \gamma_k$$

$$\rho_k = \frac{\gamma_k}{\gamma_0}$$

$$\text{Con } \gamma_0 = \sigma_t^2$$

Citando Hamilton (1994) [12], “Un proceso es estrictamente estacionario si sus propiedades no se ven afectadas por un cambio en el período de análisis, si la distribución conjunta asociada a un conjunto de m observaciones X_1, X_2, \dots, X_m es la misma a la asociada a $X_{1+k}, X_{2+k}, \dots, X_{m+k}$.

La serie de tiempo $\{X_t, t \in Z\}$ se dice estrictamente estacionaria si las distribuciones conjuntas de $(X_{t_1}, \dots, X_{t_k})'$ y $(X_{t_1+h}, \dots, X_{t_k+h})'$ son las mismas para todo entero positivo k , y para todo $t_1, \dots, t_k, k \in Z$. La distribución conjunta de cualquier conjunto de variables no se modifica si trasladamos las variables en el tiempo.”

En la realidad económica la gran mayoría de las series de tiempo son no estacionarias, por lo tanto es necesario transformarlas para lograr que la serie transformada sea estacionaria y luego ajustar algún modelo, en este caso ARIMA estacional multiplicativo con intervenciones para hacer predicciones.

Una serie no estacionaria es aquella que incumple alguno de los supuestos enunciados.

A.1.3. Metodología aplicada para la identificación del modelo

Como se mencionó anteriormente la metodología Box - Jenkins permite la estimación del modelo que ajuste para explicar la evolución de la serie estudiada a lo largo del

tiempo. Se trata de un procedimiento iterativo de prueba y error, cuyo objetivo es encontrar el modelo que mejor se ajuste a la serie.

Esta metodología sigue un proceso que consta de cuatro fases:

1. Identificación y transformaciones estacionarias
2. Estimación
3. Diagnóstico y Selección
4. Predicción

Identificación y transformaciones estacionarias

Según Hamilton, la identificación se realiza por medio de las funciones de autocorrelación simple y autocorrelación parcial.

La función de autocorrelación simple (ACF) es la representación de los coeficientes de autocorrelación del proceso en función de los distintos retardos.

La función de autocorrelación parcial (PACF) incluye los coeficientes de correlación entre observaciones separadas por k períodos cuando se elimina la dependencia producida por los valores intermedios.

En esta etapa el objetivo consiste en deducir, a partir de los datos, utilizando la ACF y la PACF muestrales, qué modelo ARIMA se adapta mejor a las características de la serie. Esto requiere:

1. Decidir qué transformaciones aplicar para convertir la serie observada en una serie estacionaria.
2. Determinar un modelo ARIMA para la serie estacionaria, es decir, los órdenes p y q de su estructura autorregresiva y de media móvil y, si el proceso es estacional, los órdenes P y Q de la estructura ARIMA estacional.

La identificación de la estructura no estacionaria consiste en detectar que transformaciones hay que aplicar para conseguir un proceso ARIMA estacionario con varianza y media constante. En particular, consiste en determinar si es necesario:

1. Transformar la serie para que tenga varianza constante

La transformación logarítmica es la transformación de los datos que cambia la escala de medición y modera el efecto de valores extremos, con el fin de corregir la heterocedasticidad.

2. Diferenciar la serie para que tenga media constante.

Para estabilizar la serie puede ser necesario aplicar diferencias regulares y estacionales. La decisión de aplicar estas diferencias puede basarse en el gráfico de la serie y en las funciones de autocorrelación muestral.

La diferenciación reemplaza cada valor de la serie por la diferencia entre este valor y el valor anterior.

La *diferenciación regular y estacional* de la serie es un filtro que se aplica cuando el componente autoregresivo muestra una raíz unitaria.

Si la serie tiene tendencia, o muestra cambios de nivel en la media, la diferenciaremos para transformarla en estacionaria.

Si la serie tiene componente estacional habrá que aplicar una diferencia estacional,

$\nabla_s = 1 - L^s$, para conseguir una serie estacionaria.

$$z_t = y_t - y_{t-1} = (1 - L)y_t \quad (\text{A.2})$$

$$w_t = y_t - y_{t-4} = (1 - L^4)y_t \quad (\text{A.3})$$

$$k_t = y_t - y_{t-12} = (1 - L^{12})y_t \quad (\text{A.4})$$

La estacionalidad se manifiesta:

- en el gráfico de la serie, que presentará una pauta repetida de período s .
- en la función de autocorreclación simple, que presentará coeficientes positivos que decrecen lentamente en los retardos $s, 2s, 3s, \dots$.

Aplicando este método tantas veces como sea necesario, ya sea en la parte regular o estacional, se conoce el valor del parámetro d y D y se logra que la serie diferenciada sea estacionaria.

Estimación

Una vez identificado el modelo de series temporales se procederá a estimar sus coeficientes²⁰, mediante máxima verosimilitud exacta.

El objetivo es la estimación de los valores de los coeficientes $((\Phi_1, \dots, \Phi_p), (\phi_1, \dots, \phi_p), (\Theta_1, \dots, \Theta_Q), (\theta_1, \dots, \theta_q))$ en base a las observaciones de la serie y_t .

Para la estimación de los parámetros se suponen las siguientes hipótesis:

- $\varepsilon_t \sim N(0, \sigma_\varepsilon), t = 1, \dots, T$.
- El proceso es estacionario.
- El proceso es invertible.

Para la estimación de los coeficientes del modelo se recurre al cálculo de estimaciones de máxima verosimilitud, es decir, estimaciones que maximicen la probabilidad de obtener los valores de la variable dependiente Y proporcionados por los datos de la serie. Estas estimaciones no son de cálculo directo sino que se recurre a métodos iterativos.

Se recuerda que si X_1, \dots, X_n son una muestra independiente, igualmente distribuidos de una población con función de distribución $f(x / \theta_1, \dots, \theta_k)$, la función de verosimilitud esta definida por

$$L(\theta / x) = L(\theta_1, \dots, \theta_k / x_1, \dots, x_n) = \prod_{i=1}^n f(x_i / \theta_1, \dots, \theta_k) \quad (\text{A.5})$$

Para cada punto muestral x , sea $\hat{\theta}(x)$ el valor del parámetro en el que $L(\theta / x)$ alcanza su valor máximo como una función de θ , con x fijo.

Si la función de verosimilitud es diferenciable (en θ_i), los posibles candidatos para estimadores máximo verosímil son los valores de $(\theta_1, \dots, \theta_k)$ que resuelven

$$\frac{\partial}{\partial \theta_i} L(\theta / x) = 0, \quad (\text{A.6})$$
$$i = 1, \dots, k$$

De estos métodos surgen no sólo las estimaciones de los coeficientes de regresión, sino también se obtienen sus errores estándar y los residuos del modelo.

²⁰ En este trabajo se utilizó el software R 2.9.2.

Diagnóstico y Selección

Una vez estimado el modelo hay que comprobar si éste se ajusta correctamente al conjunto de datos, si el modelo no resultase válido, sería necesario reformularlo.

Para la aceptación del modelo deben analizarse los coeficientes estimados, la bondad de ajuste del modelo y los residuos.

- Para la significación de cada uno de los coeficientes del modelo se utiliza la prueba de hipótesis (prueba-t):

$$H_0) \beta_j = 0$$

$$H_a) \beta_j \neq 0$$

Cuyo estadístico de prueba es:

$$t_{(obs)}(j) = \frac{\hat{\beta}_j}{\sqrt{Var(\hat{\beta}_j)}} \square N(0, \sigma_\beta^2) \quad (A.7)$$

Cuya distribución es Normal (si contamos con un número grande de observaciones)

Se rechazará H_0 si $|t_j| \geq t^*$ o lo que es equivalente p-valor $\geq \alpha$

Tomando $\alpha = 5\%$ y $t^* = 1,96$ en el percentil 95.

En consecuencia, si alguno de ellos no fuera significativo, habría que eliminarlo del modelo, reduciendo el orden del modelo o, en su caso, prescindiendo del término independiente.

- Para las pruebas de ajuste se emplea el criterio AIC (Akaike Information Criterion) que es de gran utilidad a la hora de comparar distintos modelos.

El *criterio AIC* es utilizado para la selección de los modelos que están considerados como posibles, para modelizar la serie de datos. El modelo consta de dos componentes: el primero $(Ln(\frac{SCR}{T}))$, referido a la minimización de los errores y el segundo $(2(\frac{1+P+Q+p+q}{T}))$, que penaliza la incorporación de parámetros adicionales.

$$AIC = Ln\left(\frac{SCR}{T}\right) + 2\left(\frac{1+P+Q+p+q}{T}\right) \quad (A.8)$$

Donde:

- SCR es la suma de cuadrados de los residuos definida según:

$$SCR = \sum_{i=1}^t (y_i - \hat{y}_i)^2 \quad (A.9)$$

- T es el número de observaciones
- p y P son el orden del polinomio autorregresivo regular y estacional respectivamente.
- q y Q son el orden del polinomio de medias móviles regular y estacional respectivamente.

El modelo con menor valor de AIC será escogido como el que mejor se ajusta a los datos. El AIC presenta una desventaja, selecciona modelos con mayor número de parámetros.

- En esta etapa además se comprueba que los residuos no tienen estructura de dependencia y siguen un proceso ruido blanco. Si los residuos no contienen información, se acepta el modelo como adecuado y se lo utiliza para predicción. Si los residuos muestran estructura, se modificará el modelo y se repetirán las tres etapas anteriores hasta obtener el modelo adecuado.

El cumplimiento de los supuestos de los residuos implica verificar los siguientes aspectos:

- Los residuos deben ser normales e incorrelacionados.

Para testear la incorrelación de los residuos se utiliza el *test de Ljung-Box*. Se considera el estadístico:

$$Q = n(n+2) \sum_{j=1}^k \rho^2(j) / (n-j) \square \chi^2(k) \quad (\text{A.10})$$

Para el cual la prueba de hipótesis planteada será:

$$H_0) \rho_i = 0, i = 1, \dots, k$$

$$H_a) \text{ al menos un } \rho_i \neq 0$$

Si el valor del estadístico Q calculado excede el valor crítico de $\chi^2(k)$, al menos algún valor de la función de autocorrelación para los primeros k retardos será significativamente distinto de cero. Lo que implica rechazar la hipótesis nula y concluir que existe autocorrelación en los primeros k retardos considerados.

- Los residuos deben tener distribución Normal.

Para testear la normalidad de los residuos se utiliza el *test de Shapiro Wilk* y el *test de Jarque Bera*.

Para corroborar si una variable se distribuye Normal se puede utilizar el estadístico W para la prueba de Shapiro-Wilk.

$$W = \frac{\sum_{i=1}^{n/2} (X_{(n-i+1)} - X_{(i)}) a_{(n-i+1)}}{(n-1)S^2} \quad (\text{A.11})$$

Donde:

- X_1, X_2, \dots, X_n es una muestra aleatoria de una variable aleatoria X.
- X_i como la i-ésima variable de menor valor: $X_1 \leq X_2 \leq \dots \leq X_{n-1} \leq X_n$.
- $a_{(n-i+1)}$ son coeficientes de Shapiro -Wilk.
- Varianza muestral:

$$S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2 \quad (\text{A.12})$$

Se acepta H_0 si $W \leq W_{n,\alpha}$, donde $W_{n,\alpha}$ es un valor de tabla para Shapiro-Wilk.

El test de Jarque-Bera se basa en la simetría y curtosis muestral para probar la normalidad y se llama así por Jarque y Bera (1980) [13].

El test de normalidad de Jarque-Bera contrasta la hipótesis nula H_0 : "los datos provienen de una distribución normal" mediante el estadístico JB que asintóticamente sigue una distribución $\chi_{(2)}$.

La hipótesis nula es una hipótesis conjunta de que la simetría es cero y el exceso de curtosis es cero, ya que la muestra de una distribución normal tiene una simetría esperada de cero y un exceso de curtosis esperado de cero (o lo que es lo mismo una curtosis de tres).

El estadístico JB es definido por:

$$JB = \frac{n}{6} \left(S^2 + \frac{(K - 3)^2}{4} \right) \quad (\text{A.13})$$

donde n es el número de observaciones, S es la simetría muestral y K es la curtosis muestral definidas como:

$$S = \frac{\hat{\mu}_3}{\hat{\sigma}^3} = \frac{\hat{\mu}_3}{(\hat{\sigma}^2)^{3/2}} = \frac{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^3}{\left(\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2 \right)^{3/2}} \quad (\text{A.14})$$

$$K = \frac{\hat{\mu}_4}{\hat{\sigma}^4} = \frac{\hat{\mu}_4}{(\hat{\sigma}^2)^2} = \frac{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^4}{\left(\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2 \right)^2} \quad (\text{A.15})$$

Donde

$\hat{\mu}_3$ y $\hat{\mu}_4$ son las estimaciones del tercer y cuarto momento central respectivamente.

\bar{x} es la media muestral.

$\hat{\sigma}_2$ es la estimación del segundo momento central, es decir, la varianza.

- Los residuos deben tener homocedasticidad de las varianzas.

El test de la homocedasticidad u homogeneidad de las varianzas de los residuos utiliza la prueba de hipótesis:

$$\begin{aligned} H_0) \sigma_1^2 = \sigma_2^2 & \iff H_0) \frac{\sigma_1^2}{\sigma_2^2} = 1 \\ H_a) \sigma_1^2 \neq \sigma_2^2 & \iff H_a) \frac{\sigma_1^2}{\sigma_2^2} \neq 1 \end{aligned}$$

La variable aleatoria F se define como el cociente de dos variables aleatorias ji-cuadrado independientes, cada una dividida entre sus respectivos grados de libertad. Esto es,

$$F = \frac{U/v_1}{V/v_2} \quad (\text{A.16})$$

donde U y V son variables aleatorias ji-cuadrado independientes con grados de libertad v_1 y v_2 respectivamente.

Se rechaza dicha hipótesis nula si la significación del estadístico F es menor que 0,05.

Predicción

Una vez que el modelo ha sido estimado y validado, está listo para ser utilizado en la predicción de valores futuros. Esta cuarta etapa consiste en realizar pruebas de performance predictiva y la predicción de los datos. En la prueba de performance predictiva se procede a cortar la serie dentro de la muestra y realizar predicciones de esos valores faltantes (predicción dentro de la muestra).

Luego se procede a realizar las predicciones fuera de la muestra.

Para evaluar las predicciones se plantea un nivel de confianza del 95%. En un nivel de confianza mayor se pierde precisión ya que los intervalos de confianza generados son más amplios pero por otro lado se gana en confianza.

Con los modelos SARIMA obtenemos valores futuros dados los valores pasados de la variable. Es deseable que la predicción puntual se aproxime lo más posible al valor observado. En el caso que existieran varios modelos candidatos, se selecciona aquel modelo que minimice el error cuadrático medio (ECM) de predicción.

$$ECM = \sum \frac{e_t^2}{N} \quad (\text{A.17})$$

RECM (Raíz del Error Cuadrático Medio)

La RECM mide la magnitud promedio del error. La diferencia entre la predicción y el correspondiente valor observado son elevados al cuadrado y luego promediados sobre la muestra. Finalmente se calcula la raíz cuadrada del promedio. Como los errores son elevados al cuadrado antes de ser promediados, la RECM da un peso relativamente

alto a los errores grandes. Esto implica que la RECM es más útil en aquellos casos en que los errores grandes no son esperados.

$$RECM = \sqrt{\sum \frac{e_t^2}{N}} \quad (A.18)$$

EAM (Error Absoluto Medio)

El EAM mide la magnitud promedio de los errores en un conjunto de predicciones, sin considerar su signo. El EAM es el promedio de los valores absolutos de las diferencias entre la predicción y la correspondiente observación. Las diferencias individuales tienen el mismo peso en el promedio.

$$EAM = \frac{1}{N} \sum |e_t| \quad (A.19)$$

Donde e_t es el error, calculado como diferencia entre los valores reales y los valores que estima el modelo.

A.1.4. Análisis de Intervención

Las series temporales pueden ser analizadas con información del pasado o ingresando información adicional. Cuando la serie se ve afectada por sucesos puntuales conocidos, como una huelga, una inundación o una crisis, la inclusión de estos efectos determinísticos en el análisis puede mejorar el resultado de la estimación del modelo y de las predicciones.

Variables impulso y escalón

Las variables ficticias más utilizadas para representar sucesos que afectan a la serie son de dos tipos: variables impulso y variables escalón. Las variables impulso representan sucesos que ocurren únicamente en un instante, por ejemplo, un accidente, un error de medida o un paro. Las variables escalón representan acontecimientos que comienzan en un instante conocido y se mantienen a partir de ese instante, por ejemplo, una subida de precios, un cambio legal, un cambio de base en un índice, etc.

Existen dos formas de modelizar la intervención:

Variable impulso

Para representar el instante de ocurrencia de este suceso vamos a definir una variable impulso por:

$$I_t^h = \begin{cases} 0, t \neq h \\ 1, t = h \end{cases} \quad (\text{A.20})$$

Variable escalón

Estas ideas pueden extenderse para modelizar intervenciones que tienen un efecto permanente sobre la serie a partir de su ocurrencia, como por ejemplo una subida de precios. Estas intervenciones se modelan con variables escalón, que se definen mediante:

$$S_t^h = \begin{cases} 0, t < h \\ 1, t \geq h \end{cases} \quad (\text{A.21})$$

La diferencia básica entre una variable impulso y una variable escalón es que el efecto de la primera se reduce a un intervalo de tiempo, mientras que el de la segunda se extiende hasta el final del período observado. Suele decirse que el efecto de un impulso en la serie es transitorio mientras que el de un escalón es permanente.

Se tiene en cuenta que un impulso es la primera diferencia de un escalón:

$$I_t^h = S_t^h - S_{t-1}^h = \nabla S_t^h \quad (\text{A.22})$$

Outliers

Según Peña (2005) [14] “Con mucha frecuencia ocurren en las series reales hechos puntuales que desconocemos. Por ejemplo, la serie puede haber estado sometida a intervenciones desconocidas como huelgas, cambios de base, errores de medición, etc. Las observaciones afectadas por estas intervenciones pueden presentar una

estructura distinta de las demás y aparecer como datos atípicos, es decir, aparentemente no generados igual que las demás. Por ejemplo, si se comete un error de medida o de transcripción en un dato, este valor no mantendrá la estructura de dependencia del resto de la serie. Es importante ser capaz de identificar estas situaciones desconocidas porque:

1. Si sus efectos son grandes pueden sesgar la estimación de los parámetros, lo que producirá a malas predicciones futuras.
2. Si el suceso ha ocurrido en la última parte de la serie y alguna observación afectada se utiliza para generar predicciones estas no serán buenas, incluso aunque los parámetros estén bien estimados.
3. Si estos sucesos atípicos pueden volver a aparecer en el futuro y los identificamos y estimamos sus efectos, podemos incorporar esta información en las predicciones y obtener intervalos de predicción más realistas.”

Una de las formas en que se pueden clasificar los outliers es en atípico aditivo (AO), cambio de nivel (LS) y cambio transitorio (TC) de acuerdo a su efecto en la serie.

Atípico Aditivo (AO)

Se considera que ha ocurrido un atípico aditivo (AO) sobre una serie temporal en el instante h si el valor de la serie se genera en ese instante de manera distinta al resto. Por ejemplo, si existe un error de medida apreciable en el instante h que no aparece en las otras observaciones, el dato z_h será un atípico aditivo en la serie z_t .

En general, si el dato z_h tiene propiedades distintas del resto por un cambio en las condiciones externas o del sistema de medición, se dice que este dato es atípico. El modelo que seguirá la serie observada, z_t si ha sido afectada por un AO en t será:

$$z_t = \begin{cases} y_t, t \neq h \\ y_t + w_A, t = h \end{cases} \quad (\text{A.23})$$

donde y_t sigue un modelo ARIMA

$$y_t = \psi(B)a_t$$

Entonces, el modelo que sigue la serie observada, z_t , es

$$z_t = w_A I_t^{(h)} + \psi(B)a_t \quad (\text{A.24})$$

Donde $I_t^{(h)} = 0, t \neq h; I_h^{(h)} = 1$.

Cambio de nivel (LS)

Un efecto importante que puede aparecer sobre una serie temporal es el cambio de nivel. Se dice que una serie ha sufrido un cambio de nivel en el instante h si sigue el modelo:

$$z_t = \omega_L S_t^h + \psi(L)a_t \quad (\text{A.25})$$

donde S_t^h es la variable escalón.

Los valores de la serie observada estarán relacionados con la serie sin contaminar por el cambio de nivel mediante:

$$z_t = \begin{cases} y_t, t < h \\ y_t + \omega_L, t \geq h \end{cases} \quad (\text{A.26})$$

Cambio transitorio (TC)

Un efecto adicional que puede considerarse es el cambio transitorio (TC), definido por:

$$z_t = \frac{\omega_{TC}}{1 - \delta L} I_t^h + \psi(L)a_t \quad (\text{A.27})$$

Se observa que si $\delta = 1$ el modelo es el del cambio de nivel, ya que $\nabla^{-1} I_t^h = S_t^h$, mientras que si $\delta = 0$ tenemos un atípico aditivo (AO).

Este tipo de atípico se utiliza fijando el valor de δ de manera que no pueda confundirse con el AO o LS. En la práctica, $\delta = 0,7$.

A.2. Metodología multivariada aplicada a los microdatos de la Encuesta de Turismo Receptivo 2008

A.2.1. Análisis de Componentes Principales (ACP)

Según plantea Blanco (2006) [15] el Análisis de Componentes Principales es una técnica factorial que se aplica a una matriz de datos del tipo individuos x variables cuantitativas.

x_{ij} es el valor de la variable j en la observación i .

I es el número de individuos (observaciones) y conjunto de ellos.

J es el número y conjunto de variables.

Cada observación esta caracterizada por una fila en el espacio R^J y cada variable es un vector en R^I formado por los valores que toman las I observaciones.

En muchos casos se les adjudica el mismo peso a los individuos, $\frac{1}{I}$ aunque hay situaciones en las que es necesario asignar pesos diferentes a distintos individuos. Estos pesos intervienen en el cálculo de las medias de cada variable y en la medida de la relación entre las variables. Siendo p_i los pesos de los individuos se tiene:

$$\bar{x}_j = \frac{1}{I} \sum_i^n x_{ij} = \sum_i^n p_i x_j \quad (\text{A.28})$$

Generalmente se le atribuye el mismo peso a las distintas variables, de no ser así sería:

$$d^2(i, l) = \sum_{j \in J} m_{ij} (x_{ij} - x_j)^2 \quad (\text{A.29})$$

El objetivo de esta técnica es obtener información esencial de la matriz de datos, reduciendo dimensiones y diferenciando los individuos que se analizan, a partir de la explicación de la mayor parte de la inercia (varianza) de los datos analizados.

La inercia, si las variables están estandarizadas es igual al número de variables.

$$\frac{1}{I} \sum_j \sum_i \left(\frac{x_{ij} - \bar{x}_j}{S_j} \right)^2 = J \quad (\text{A.30})$$

Se definen dos nubes: la nube de las filas, donde cada individuo se representa por un punto en el espacio de las variables y la nube de las columnas, donde cada variable define un punto en el espacio de los individuos.

La Nube de los Individuos (N_I)

La matriz es observada como un conjunto de filas, donde cada individuo está representado por un punto de un espacio vectorial R^J . La similitud entre los individuos está dada por la distancia euclídea entre ellos. El conjunto de distancias entre los individuos forma la nube N_I .

La Nube de las Variables (N_J)

La matriz es observada como un conjunto de columnas, cada variable está representada por un punto en el espacio R^I donde cada dimensión representa un individuo. Elegir la distancia en R^I consiste en asignar a cada dimensión un coeficiente igual a los pesos de cada individuo en la nube N_I de R^J .

También interesa determinar la distancia entre los individuos y los ángulos que forman las variables (correlación).

La distancia entre dos individuos i, l será:

$$d^2(i, l) = \sum_{j \in J} (x_{ij} - x_{lj})^2 \quad (\text{A.31})$$

La relación entre dos variables se mide mediante el coeficiente de correlación lineal entre ellas:

$$r_{(j,h)} = \frac{\text{cov}(j, h)}{\sqrt{\text{var}(j) \text{var}(h)}} = \frac{1}{I} \sum_{i \in I} \left(\frac{x_{ij} - \bar{x}_j}{S_j} \right) \left(\frac{x_{ih} - \bar{x}_h}{S_h} \right) \quad (\text{A.32})$$

La interpretación del coeficiente de correlación como un coseno es una propiedad muy importante ya que aporta un elemento geométrico al coeficiente de correlación. Las coordenadas de las variables originales proyectadas en el espacio de los individuos coinciden con su coeficiente de correlación.

A dichos efectos, se realiza un procedimiento factorial en cada nube - proyección en un subespacio menor o igual que el original - encontrando ejes de inercia tales que la inercia de la nube proyectada sobre esos ejes sea máxima. El conjunto de las coordenadas de los puntos originales de la nube proyectada sobre cada uno de estos nuevos ejes define una combinación lineal llamada factor (o componente). Los factores (F_s) así definidos se llaman Componentes Principales de la nube de individuos.

Las coordenadas de los individuos en los nuevos ejes serán: $F_s = XM u_s$ siendo X la matriz de datos estandarizada. Siendo M matriz de los pesos de las columnas.

Las coordenadas de las variables proyectadas en los nuevos ejes serán: $G_s = X' D v_s$, con X la matriz de datos estandarizada. Siendo D una matriz diagonal, matriz de pesos de las filas.

Existe una relación de dualidad entre las proyecciones de filas y columnas. Cada una de las nubes proyectadas sobre un conjunto de ejes ortogonales produce una descomposición de la inercia total y aquellas inercias de mismo rango son iguales.

A efectos de determinar la calidad de los subespacios de proyección en base a la información que poseen de la nube original, se utilizan ayudas a la interpretación. Los índices más tradicionales se describen a continuación.

- Calidad de representación global de una nube por un subespacio: Es el porcentaje de inercia explicada por los sucesivos espacios de proyección en relación a la inercia total.
- Calidad de representación de un elemento: Es el cociente entre la inercia de la proyección del elemento i sobre el eje s y la inercia total de i , que es el coseno cuadrado del ángulo que forma el vector original y el eje s .
- Contribución de un elemento a la inercia de un eje: Es útil para medir la porción de la inercia del eje s explicada por una variable o conjunto de variables.

Una vez seleccionado el espacio de proyección, se caracterizan los ejes a partir de las variables originales con mayor correlación con los componentes principales.

A.2.2. Análisis de Clusters

Según plantea Blanco (2006), el Análisis de Clusters o de conglomerados es una técnica de clasificación aplicable al ámbito de la información multivariada, que busca clasificar observaciones en distintos grupos, de acuerdo a ciertas características estudiadas. Básicamente, se trata de formar grupos de manera tal que aquellos objetos en el mismo grupo sean similares entre sí, mientras objetos en diferentes grupos sean tan disímiles como sea posible.

Para ello es necesario contar con información de individuos expresada en variables cuantitativas, definir un criterio de similitud, de asociación de individuos, y de cuándo detenerse en la formación de grupos (es decir, con cuántos grupos quedarse).

Las etapas de aplicación de un método de clasificación son:

- a) definición de la matriz de datos originales
- b) determinación de un concepto de similitud o disimilaridad entre observaciones
- c) definición de algoritmos de similaridad
- d) chequeo de reglas de detención (stopping rules)
- e) selección del número de grupos y descripción de cada uno de ellos en función de las características observadas

En el presente estudio se utiliza la distancia euclidiana, que para dos individuos i y j se define de la siguiente forma:

$$d_{ij}^2 = \sum_{k=1}^p (x_{ik} - x_{jk})^2 \quad (\text{A.33})$$

donde:

p = número de variables consideradas

x_{ik} = el valor del individuo i en la variable k

Los métodos de clasificación pueden dividirse en agregativos o divisivos. A su vez, pueden clasificarse en jerárquicos y no jerárquicos. El procedimiento de los métodos agregativos jerárquicos parte de que cada individuo es un grupo, por lo que en cada paso se irán uniendo individuos con individuos, o individuos con grupos, o grupos con grupos.

En cada paso se forman particiones encajadas, por lo tanto un grupo formado en un paso no se desarma en los siguientes pasos. El método termina con la unión de todos los individuos en un solo grupo. Por lo tanto, la cantidad de grupos no se define a priori, sino que se lleva a cabo utilizando un conjunto de indicadores como R^2 , pseudo F y pseudo t^2 .

Las técnicas de análisis de clusters pueden resultar de mayor utilidad cuando se aplican a valores derivados de los datos que buscan resumir los datos originales en una forma óptima. Uno de los métodos más comunes y también de más utilidad es el análisis de componentes principales.

Las clasificaciones jerárquicas pueden ser representadas por un diagrama en dos dimensiones conocido como dendrograma, que permite reconstruir la historia de la formación de los clusters, a partir de las fusiones realizadas en cada etapa del análisis.

Los algoritmos de clasificación de tipo agregativo más utilizados son:

- Técnica del vecino más lejano: Bajo este método, se parte de los I individuos, cada uno considerado un grupo. Con una distancia d_{ij} se evalúa la proximidad de los objetos a clasificar. La estrategia de este método consiste en elegir como distancia entre la clase $\langle i, j \rangle$ y el objeto k la distancia más grande entre $d_{(i,k)}$ y $d_{(j,k)} \forall k \in I$. En cada paso, una vez construida la matriz de distancias según la distancia máxima, se unen aquellas clases (sean formadas por varios individuos o por un solo individuo) que tienen la mínima distancia entre ellas, es decir aquellas que son más próximas.
- Método de Ward: Procedimiento de clustering que busca formar las particiones minimizando la pérdida de información (definida en términos de un criterio de suma de los cuadrados de los errores) asociada con cada cluster. En cada instancia, este método minimiza la variabilidad dentro de los grupos y maximiza la variabilidad entre grupos.

En el presente trabajo se utiliza como criterio de asociación el método de Ward, que busca optimizar en cada etapa la dispersión (o varianza) de las clases de la partición obtenida por agregación de dos objetos. La varianza total se puede descomponer en la varianza dentro de los grupos (within) y la varianza entre grupos (between).

$$\sum_{i=1}^I \sum_{j=1}^J (x_{ij} - x_j)^2 = \sum_{k=1}^K \sum_{i=1}^{n_k} \sum_{j=1}^J (x_{ij(k)} - \bar{x}_{kj})^2 + \sum_{k=1}^K \sum_{j=1}^J n_k (\bar{x}_{kj} - \bar{x}_j)^2 \quad (\text{A.34})$$

Varianza total = Varianza Intra-grupo (Within) + Varianza Entre grupos (Between)

Cada vez que se unen dos grupos, el grupo resultante será más heterogéneo que cada grupo por separado, porque se unen elementos que tienen algún tipo de diferencia. El método consiste en minimizar el crecimiento de la variación intra-grupos resultante de la agregación de dos grupos en una nueva clase.

Los indicadores de detención utilizados son:

- pseudo F: Busca máximos relativos en relación a la varianza entre clusters (between) con respecto a la varianza dentro de éstos (within).

$$F = \frac{\sum \sum n_k (\bar{x}_{kj} - \bar{x}_j)^2 / (k-1)}{\sum \sum (x_{ij} - \bar{x}_j)^2 / (n-k)} \quad (\text{A.35})$$

- pseudo t²: Si del paso k al k+1 se produce una caída importante del indicador, se considera significativa la fusión de grupos (u observaciones) en k+1.

$$t^2 = \frac{tr(W_{GL}) - (tr(W_G) + tr(W_L))}{(tr(W_G) + tr(W_L)) / (n_G + n_L - 2)} \quad (\text{A.36})$$

Siendo W la matriz de varianzas y covarianzas y $tr(W) = \sum \text{varianzas}$.

- R²: Establece la relación entre la variación explicada y la variación total, buscando maximizar su valor.

$$R^2 = \frac{\sum \sum n_k (\bar{x}_{kj} - \bar{x}_j)^2}{\sum \sum (x_{ij} - \bar{x}_j)^2}, \quad (\text{A.37})$$

cuando existen N grupos $R^2 = 1$, cuando hay un solo grupo $R^2 = 0$.

Como plantea Blanco (2006), “A diferencia de los métodos jerárquicos donde la agrupación se realiza mediante un proceso de agrupación (desagrupación) sucesiva cuyo resultado final es una jerarquía de unión completa en la que cada grupo se une (separa) en una determinada fase, en los métodos no jerárquicos se produce una partición en un predeterminado número de clases que no se solapan y que, por ejemplo, en algún método permiten la reasignación de elementos. Algunos de los métodos no-jerárquicos más conocidos son clasificación alrededor de centros móviles y k-medias que requieren de la especificación previa del número de grupos.”

Este trabajo considera solamente el método k-medias (2010) [16], que es el que se utiliza. El método de k-medias permite procesar un número ilimitado de individuos pero solo permite utilizar un método agregativo y requiere que se proponga previamente el número de clusters que se desea obtener. El análisis de clusters de k-medias es un método de agregación de individuos que se basa en las distancias existentes entre ellos. Se comienza seleccionando los k individuos más distantes entre sí (se debe determinar inicialmente el número k de clusters que se desea obtener) y a continuación se asigna cada individuo al centro más próximo y se actualiza el valor de los centros a medida que se van incorporando nuevos individuos. Una vez que todos los individuos han sido asignados a uno de los k clusters, se inicia un proceso iterativo para calcular los centroides finales de esos k clusters.

El algoritmo de las k-medias es un proceso iterativo de refinamiento de un número k de clusters, definidos a priori. La necesidad de proporcionar al algoritmo este número de grupos a identificar hace que el analista deba estimar dentro del problema a tratar el número de sub-poblaciones que se van a intentar identificar, dato que no necesariamente se puede conocer. El procedimiento del algoritmo de k-medias consta de los siguientes pasos:

1. El algoritmo comienza con la selección aleatoria de k posiciones aleatorias del espacio de datos.
2. Para cada uno de los elementos de los datos de entrada se calcula el punto más próximo de los k elegidos. Esta asignación determina el cluster al que pertenece cada dato.
3. Se recalcula para cada uno de los clusters el punto medio. Y se repite el proceso del paso 2.
4. El algoritmo concluye cuando los clusters permanecen incambiados y no se han modificado los centros medios a lo largo de dos iteraciones.

Las restricciones de este método son:

- La necesidad de estimar el número k de clusters a identificar al comienzo del algoritmo.
- En cada iteración se tienen que identificar los nuevos puntos medios de cada cluster. Para datos continuos o numéricos esto es factible, pero en datos categóricos y discretos no existe una función para determinar dichos valores medios. En estos casos se usa una variante de este algoritmo denominado k-

medioides. En esta variante no se identifican los puntos medios de cada cluster sino el objeto más próximo a dicho punto central.

Como plantean Ouellette *et al.* (2010) [17] “En el método k-medias se utiliza el índice “Calinski” para determinar el número óptimo de clusters. Se refiere al conocido criterio desarrollado por Calinski-Harabasz (1974) [18].

Regla de Calinski:

$$(SSB / (K - 1)) / (SSW / (n - K)), \quad (A.38)$$

donde n es el número de observaciones y K es el número de clusters. SSW es la suma de cuadrados dentro de los clusters mientras que SSB es la suma de cuadrados entre los clusters.

En un estudio de simulación, Milligan y Cooper (1985) [19] encontraron que el criterio Calinski-Harabasz proporcionaba el número correcto de grupos la mayoría de las veces. Milligan y Cooper recomiendan este criterio porque el máximo valor de “Calinski” generalmente indica el número correcto de grupos”.

APÉNDICE B: DATOS

B.1. Ingreso trimestral de turistas a Uruguay (1993.01-2009.01)

Trimestre	Nro. De Turistas	2001-1T	975144
1993-1T	842802	2001-2T	369804
1993-2T	320984	2001-3T	335494
1993-3T	334355	2001-4T	456004
1993-4T	455272	2002-1T	533417
1994-1T	881595	2002-2T	233156
1994-2T	337938	2002-3T	249832
1994-3T	366016	2002-4T	337467
1994-4T	589918	2003-1T	518737
1995-1T	811511	2003-2T	251884
1995-2T	388053	2003-3T	285007
1995-3T	425947	2003-4T	452427
1995-4T	551419	2004-1T	726848
1996-1T	853604	2004-2T	310391
1996-2T	417545	2004-3T	318042
1996-3T	409123	2004-4T	515577
1996-4T	578344	2005-1T	777758
1997-1T	997556	2005-2T	273018
1997-2T	402614	2005-3T	340419
1997-3T	452998	2005-4T	525854
1997-4T	609364	2006-1T	694292
1998-1T	867390	2006-2T	305994
1998-2T	440975	2006-3T	330609
1998-3T	409848	2006-4T	493445
1998-4T	605780	2007-1T	625461
1999-1T	938636	2007-2T	305935
1999-2T	375442	2007-3T	341585
1999-3T	390139	2007-4T	542300
1999-4T	568947	2008-1T	733440
2000-1T	999907	2008-2T	317096
2000-2T	350282	2008-3T	380560
2000-3T	348562	2008-4T	566788
2000-4T	537136	2009-1T	726441

B.2. Ingreso mensual de turistas a Uruguay (1996.01 – 2009.07)

Mes	Nro. de Turistas	200007	102508	200502	231946
199601	311350	200008	88903	200503	156521
199602	281761	200009	86181	200504	85197
199603	150844	200010	110265	200505	78208
199604	146541	200011	111653	200506	68768
199605	96169	200012	214649	200507	108308
199606	99932	200101	381091	200508	83867
199607	120556	200102	308713	200509	89738
199608	112054	200103	152439	200510	113898
199609	103978	200104	146340	200511	117205
199610	120536	200105	76795	200512	196837
199611	143332	200106	68187	200601	289151
199612	205532	200107	107633	200602	203283
199701	363922	200108	86927	200603	105188
199702	291333	200109	76205	200604	105139
199703	193127	200110	100839	200605	72434
199704	121130	200111	109376	200606	68728
199705	119318	200112	174174	200607	101285
199706	81648	200201	197958	200608	86975
199707	132537	200202	146416	200609	95064
199708	119464	200203	108489	200610	110714
199709	108273	200204	59293	200611	118176
199710	119455	200205	67840	200612	193694
199711	142312	200206	56525	200701	247357
199712	227478	200207	77705	200702	197181
199801	345016	200208	55607	200703	118092
199802	258129	200209	53780	200704	106579
199803	144409	200210	67242	200705	81283
199804	156454	200211	72194	200706	75279
199805	108862	200212	132323	200707	109715
199806	93138	200301	192471	200708	82922
199807	126040	200302	156410	200709	98593
199808	104358	200303	90347	200710	111387
199809	98438	200304	87229	200711	132379
199810	144095	200305	57469	200712	219586
199811	122550	200306	57745	200801	286482
199812	232846	200307	90975	200802	225299
199901	387216	200308	69927	200803	147872
199902	270937	200309	70351	200804	92725
199903	156493	200310	94796	200805	95492
199904	126972	200311	90884	200806	79941
199905	92391	200312	184634	200807	122280
199906	84947	200401	300316	200808	108215
199907	123382	200402	234254	200809	97549
199908	94117	200403	104167		

199909	94186	200404	123431	200810	126139
199910	127710	200405	66195	200811	130551
199911	125524	200406	62268	200812	231510
199912	204429	200407	105466	200901	281925
200001	393903	200408	77287	200902	223561
200002	309370	200409	81540	200903	139164
200003	149874	200410	113513	200904	140922
200004	134399	200411	107457	200905	110788
200005	72119	200412	197689	200906	86307
200006	71497	200501	307726	200907	106016

APÉNDICE C: RESULTADOS DEL MODELO 1

C.1. Cumplimiento de los supuestos de los residuos para el modelo seleccionado: ARIMA(1,0,0)(0,1,0)(4) con Turismo y Tc2002

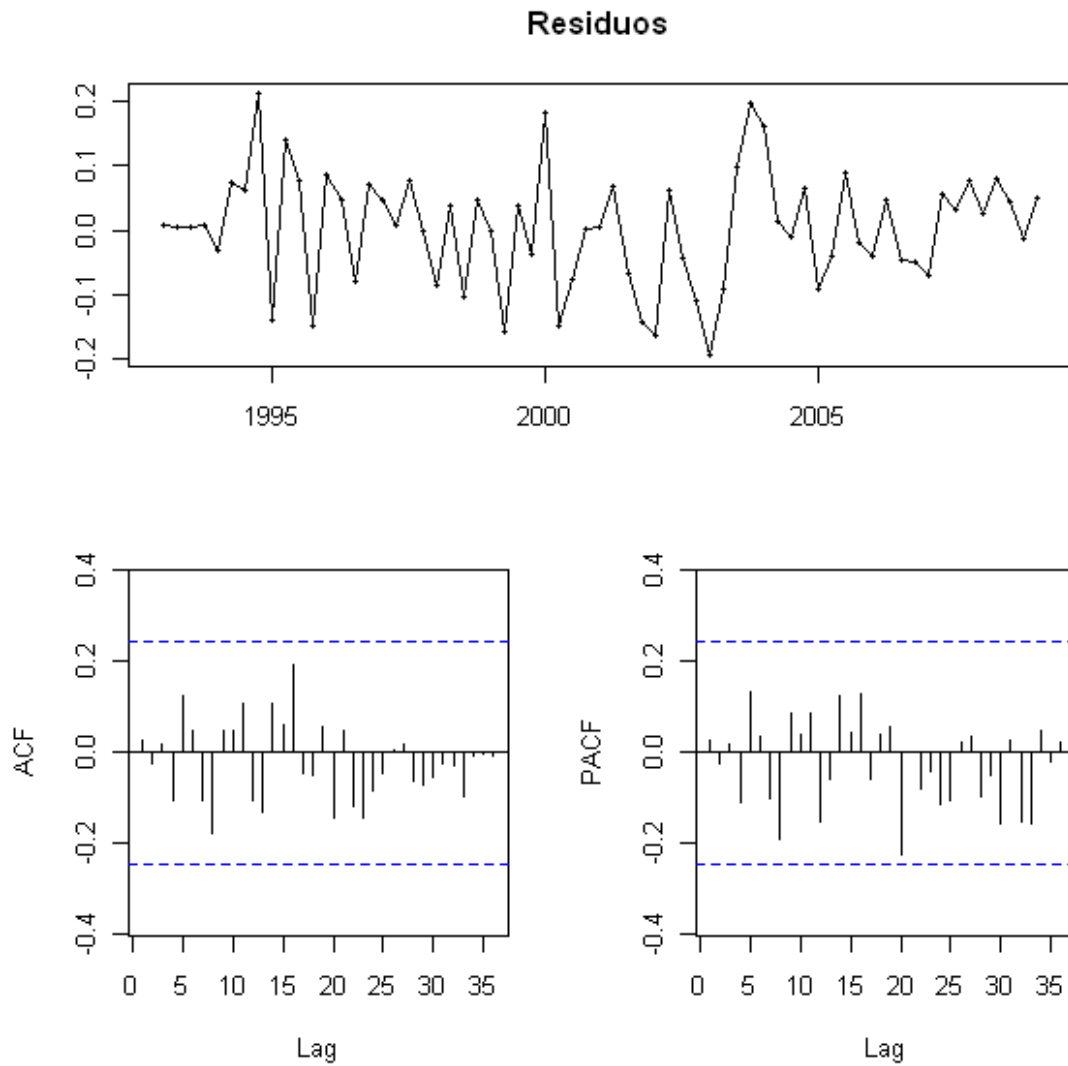


Figura C.1. Gráfico y autocorrelograma simple y parcial de los residuos para el modelo 1.

Media de los residuos	0,003
Desvío estándar de los residuos	0,089
Varianza de los residuos	0,007

Cuadro C.1. Medidas de los residuos.

Ljung-Box	df = 4, p-valor = 0,92
Ljung-Box	df = 8, p-valor = 0,71
Ljung-Box	df = 12, p-valor = 0,81
Ljung-Box	df = 16, p-valor = 0,62

Cuadro C.2. Test de Ljung-Box.

Shapiro-Wilk	p-valor = 0,37
Jarque-Bera	p-valor = 0,92

Cuadro C.3. Test de Shapiro-Wilk y Jarque-Bera.

C.2. Modelo descartado: ARIMA(1,0,0)(0,1,0)(4) con Turismo y Ls2002

ARIMA (1,0,0) (0,1,0) ₍₄₎	Coefficientes	t observados
Ar1	0,53	4,95
Turismo	0,07	4,46
Ls2002	-0,42	-6,28
AIC	-108,48	

Cuadro C.4. Modelo descartado.

C.3. Comparación de modelos

	Modelo Seleccionado: ARIMA (1,0,0) (0,1,0) ₍₄₎ con Turismo y Tc2002	Modelo Descartado: ARIMA (1,0,0) (0,1,0) ₍₄₎ con Turismo y Ls2002
RECM	0,090	0,092
EAM	0,071	0,072

Cuadro C.5. Comparación de modelos.

C.4. Predicción dentro de la muestra actualizando a un paso para el modelo seleccionado: ARIMA(1,0,0)(0,1,0)(4) con Turismo y Tc2002

Trimestre	Predicción	Valor Real	Int. Inf. 95	Int. Sup. 95	Error	Error en %
2T-2008	290.302	317.096	241.675	348.714	26.794	8,4%
3T-2008	348.137	380.560	283.166	428.015	32.423	8,5%
4T-2008	547.689	566.788	442.888	677.288	19.099	3,4%
1T-2009	674.461	726.441	544.562	835.347	51.980	7,2%

Cuadro C.6. Predicción dentro de la muestra actualizando a un paso.

Trimestre	Predicción	Valor Real	Int. Inf. 95	Int. Sup. 95	Error	Error en %
3T-2008	363.936	380.560	303.070	437.026	16.624	4,4%
4T-2008	560.744	566.788	455.941	689.637	6.044	1,1%
1T-2009	686.148	726.441	554.477	849.086	40.293	5,5%

Cuadro C.7. Predicción dentro de la muestra actualizando a un paso.

Trimestre	Predicción	Valor Real	Int. Inf. 95	Int. Sup. 95	Error	Error en %
4T-2008	574.603	566.788	479.093	689.153	-7.815	-1,4%
1T-2009	694.765	726.441	565.314	853.858	31.676	4,4%

Cuadro C.8. Predicción dentro de la muestra actualizando a un paso.

Trimestre	Predicción	Valor Real	Int. Inf. 95	Int. Sup. 95	Error	Error en %
1T-2009	689.631	726.441	575.855	825.886	36.810	5,1%

Cuadro C.9. Predicción dentro de la muestra actualizando a un paso.

APÉNDICE D: RESULTADOS DEL MODELO 2

D.1. Cumplimiento de los supuestos de los residuos para el modelo seleccionado: ARIMA(1,0,0)(0,1,1)(12) con Turismo, Tc Enero 2002 y Tc Mayo 2002

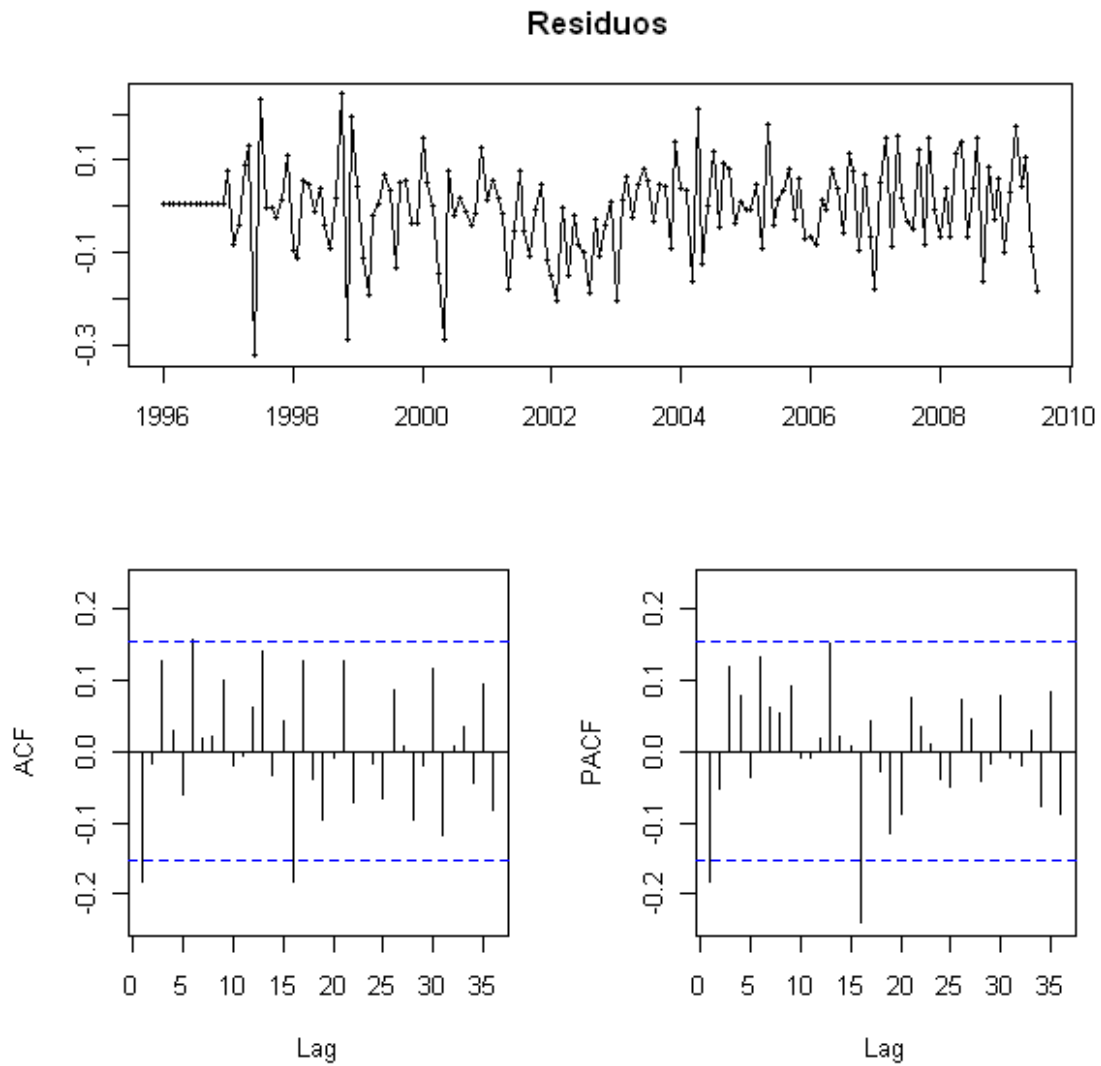


Figura D.1. Gráfico y autocorrelograma simple y parcial de los residuos para el modelo 2.

Media de los residuos	-0,003
Desvío estándar de los residuos	0,098
Varianza de los residuos	0,009

Cuadro D.1. Medidas de los residuos.

Ljung-Box	df = 12, p-valor = 0,19
Ljung-Box	df = 24, p-valor = 0,06
Ljung-Box	df = 36, p-valor = 0,07
Ljung-Box	df = 48, p-valor = 0,17

Cuadro D.2. Test de Ljung-Box.

Shapiro-Wilk	p-valor = 0,057
--------------	-----------------

Cuadro D.3. Test de Shapiro-Wilk.

D.2. Modelo descartado: ARIMA(1,0,0)(1,1,0)(12) con Turismo, Tc Enero 2002 y Tc Mayo 2002

ARIMA (1,0,0) (1,1,0) ₍₁₂₎	Coefficientes	t observados
Ar1	0,83	17,67
Sar1	-0,42	-5,04
Turismo	0,26	13,40
Tc Enero 2002	-0,36	-4,16
Tc Mayo 2002	0,39	4,40
AIC	-232,6	

Cuadro D.4. Modelo descartado.

D.3. Comparación de modelos

	Modelo Seleccionado: ARIMA (1,0,0) (0,1,1) ₍₁₂₎ con Turismo, Tc Enero 2002 y Tc Mayo 2002	Modelo Descartado: ARIMA (1,0,0) (1,1,0) ₍₁₂₎ con Turismo, Tc Enero 2002 y Tc Mayo 2002
RECM	0,097	0,102
EAM	0,072	0,077

Cuadro D.5. Comparación de modelos.

D.4. Predicción dentro de la muestra actualizando a un paso para el modelo seleccionado: ARIMA(1,0,0)(0,1,1)(12) con Turismo, Tc Enero 2002 y Tc Mayo 2002

Mes	Predicción	Valor Real	Int. Inf. 95	Int. Sup. 95	Error	Error en %
set-08	114.811	97.549	94.123	140.047	-17.262	-17,7%
oct-08	133.637	126.139	102.860	173.622	-7.498	-5,9%
nov-08	141.180	130.551	104.610	190.533	-10.629	-8,1%
dic-08	233.908	231.510	169.015	323.718	-2.398	-1,0%
ene-09	316.204	281.925	224.555	445.257	-34.279	-12,2%
feb-09	239.456	223.561	167.989	341.325	-15.895	-7,1%
mar-09	123.665	139.164	86.004	177.818	15.499	11,1%
abr-09	122.941	140.922	84.965	177.890	17.981	12,8%
may-09	88.220	110.788	60.692	128.232	22.568	20,4%
jun-09	77.866	86.307	53.392	113.558	8.441	9,8%
jul-09	116.924	106.016	79.981	170.933	-10.908	-10,3%

Cuadro D.6. Predicción dentro de la muestra actualizando a un paso.

Mes	Predicción	Valor Real	Int. Inf. 95	Int. Sup. 95	Error	Error en %
oct-08	116.742	126.139	95.671	142.455	9.397	7,4%
nov-08	124.809	130.551	96.095	162.102	5.742	4,4%
dic-08	210.811	231.510	156.385	284.177	20.699	8,9%
ene-09	291.748	281.925	211.232	402.957	-9.823	-3,5%
feb-09	223.192	223.561	158.945	313.406	0.369	0,2%
mar-09	116.258	139.164	81.846	165.138	22.906	16,5%
abr-09	116.408	140.922	81.290	166.698	24.514	17,4%
may-09	83.282	110.788	57.823	119.950	27.506	24,8%
jun-09	74.228	86.307	51.324	107.352	12.079	14,0%
jul-09	112.125	106.016	77.299	162.642	-6.109	-5,8%

Cuadro D.7. Predicción dentro de la muestra actualizando a un paso.

Mes	Predicción	Valor Real	Int. Inf. 95	Int. Sup. 95	Error	Error en %
nov-08	133.060	130.551	109.102	162.278	-2.509	-1,9%
dic-08	222.836	231.510	171.643	289.297	8.674	3,7%
ene-09	306.924	281.925	227.728	413.661	-24.999	-8,9%
feb-09	233.036	223.561	168.711	321.887	-9.475	-4,2%
mar-09	120.588	139.164	85.848	169.385	18.576	13,3%
abr-09	120.048	140.922	84.467	170.616	20.874	14,8%
may-09	85.198	110.788	59.451	122.097	25.590	23,1%
jun-09	75.727	86.307	52.529	109.171	10.580	12,3%
jul-09	114.028	106.016	78.759	165.090	-8.012	-7,6%

Cuadro D.8. Predicción dentro de la muestra actualizando a un paso.

Mes	Predicción	Valor Real	Int. Inf. 95	Int. Sup. 95	Error	Error en %
dic-08	219.337	231.510	179.992	267.284	12.173	5,3%
ene-09	303.396	281.925	233.932	393.486	-21.471	-7,6%
feb-09	230.729	223.561	171.383	310.625	-7.168	-3,2%
mar-09	119.536	139.164	86.640	164.924	19.628	14,1%
abr-09	119.140	140.922	84.916	167.157	21.782	15,5%
may-09	84.472	110.788	59.504	119.915	26.316	23,8%
jun-09	75.199	86.307	52.533	107.643	11.108	12,9%
jul-09	113.329	106.016	78.701	163.192	-7.313	-6,9%

Cuadro D.9. Predicción dentro de la muestra actualizando a un paso.

Mes	Predicción	Valor Real	Int. Inf. 95	Int. Sup. 95	Error	Error en %
ene-09	318.250	281.925	261.316	387.587	-36.325	-12,9%
feb-09	240.383	223.561	185.450	311.589	-16.822	-7,5%
mar-09	123.798	139.164	91.994	166.598	15.366	11,0%
abr-09	122.760	140.922	88.997	169.333	18.162	12,9%
may-09	86.541	110.788	61.684	121.413	24.247	21,9%
jun-09	76.795	86.307	54.090	109.031	9.512	11,0%
jul-09	115.376	106.016	80.579	165.201	-9.360	-8,8%

Cuadro D.10. Predicción dentro de la muestra actualizando a un paso.

Mes	Predicción	Valor Real	Int. Inf. 95	Int. Sup. 95	Error	Error en %
feb-09	216.822	223.561	177.877	264.292	6.739	3,0%
mar-09	113.462	139.164	87.509	147.112	25.702	18,5%
abr-09	114.061	140.922	84.805	153.409	26.861	19,1%
may-09	82.093	110.788	59.600	113.076	28.695	25,9%
jun-09	73.292	86.307	52.358	102.595	13.015	15,1%
jul-09	110.944	106.016	78.375	157.048	-4.928	-4,6%

Cuadro D.11. Predicción dentro de la muestra actualizando a un paso.

Mes	Predicción	Valor Real	Int. Inf. 95	Int. Sup. 95	Error	Error en %
mar-09	116.450	139.164	95.593	141.858	22.714	16,3%
abr-09	116.606	140.922	90.006	151.069	24.316	17,3%
may-09	83.644	110.788	62.246	112.398	27.144	24,5%
jun-09	74.466	86.307	54.114	102.473	11.841	13,7%
jul-09	112.452	106.016	80.413	157.257	-6.436	-6,1%

Cuadro D.12. Predicción dentro de la muestra actualizando a un paso.

Mes	Predicción	Valor Real	Int. Inf. 95	Int. Sup. 95	Error	Error en %
abr-09	135.185	140.922	110.805	164.928	5.737	4,1%
may-09	95.198	110.788	73.326	123.594	15.590	14,1%
jun-09	83.106	86.307	61.688	111.958	3.201	3,7%
jul-09	123.475	106.016	89.472	170.401	-17.459	-16,5%

Cuadro D.13. Predicción dentro de la muestra actualizando a un paso.

Mes	Predicción	Valor Real	Int. Inf. 95	Int. Sup. 95	Error	Error en %
may-09	98.584	110.788	80.851	120.206	12.204	11,0%
jun-09	85.638	86.307	65.999	111.122	0.669	0,8%
jul-09	126.700	106.016	94.087	170.617	-20.684	-19,5%

Cuadro D.14. Predicción dentro de la muestra actualizando a un paso.

Mes	Predicción	Valor Real	Int. Inf. 95	Int. Sup. 95	Error	Error en %
jun-09	94.529	86.307	77.481	115.328	-8.222	-9,5%
jul-09	137.964	106.016	106.217	179.200	-31.948	-30,1%

Cuadro D.15. Predicción dentro de la muestra actualizando a un paso.

Mes	Predicción	Valor Real	Int. Inf. 95	Int. Sup. 95	Error	Error en %
jul-09	127.544	106.016	104.546	155.602	-21.528	-20,3%

Cuadro D.16. Predicción dentro de la muestra actualizando a un paso.

APÉNDICE E: RESULTADOS DEL MODELO 3

E.1. Cumplimiento de los supuestos de los residuos para el modelo seleccionado: ARIMA(0,0,1)(0,1,1)(4) con Ls2002

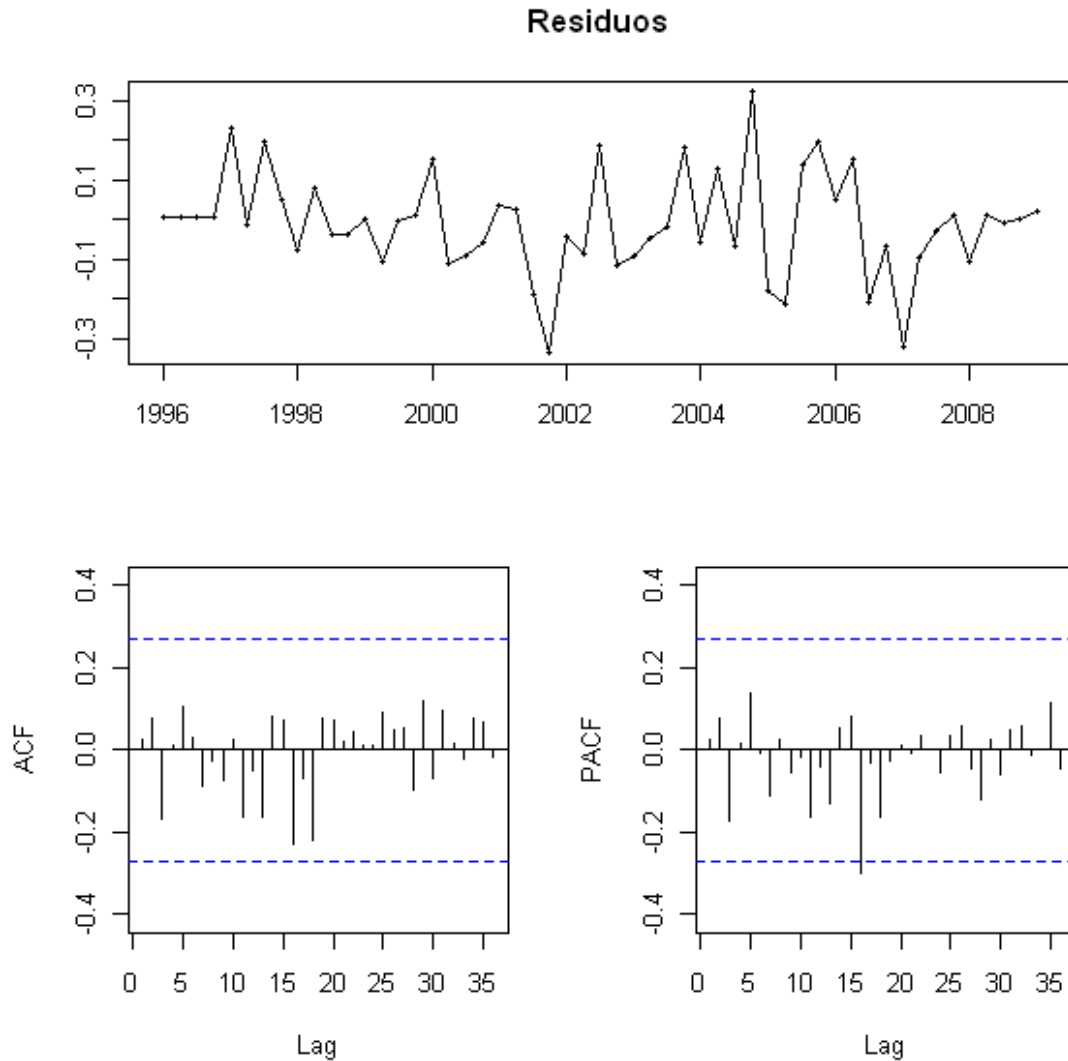


Figura E.1. Gráfico y autocorrelograma simple y parcial de los residuos para el modelo 3.

Media de los residuos	-0,011
Desvío estándar de los residuos	0,130
Varianza de los residuos	0,016

Cuadro E.1. Medidas de los residuos.

Ljung-Box	df = 4, p-valor = 0,73
Ljung-Box	df = 8, p-valor = 0,91
Ljung-Box	df = 12, p-valor = 0,93
Ljung-Box	df = 16, p-valor = 0,70

Cuadro E.2. Test de Ljung-Box.

Shapiro-Wilk	p-valor = 0,19
Jarque-Bera	p-valor = 0,74

Cuadro E.3. Test de Shapiro-Wilk y Jarque-Bera.

E.2. Modelo descartado: ARIMA(1,0,0)(0,1,1)₍₄₎ con Ls2002

ARIMA (1,0,0) (0,1,1) ₍₄₎	Coefficientes	t observados
Ar1	0,35	2,36
Sma1	-0,77	-5,21
Ls2002	-0,31	-3,81
AIC	-45,79	

Cuadro E.4. Modelo descartado.

E.3. Predicción dentro de la muestra actualizando a un paso para el modelo seleccionado: ARIMA(0,0,1)(0,1,1)(4) con Ls2002

Trimestre	Predicción	Valor Real	Int. Inf. 95	Int. Sup. 95	Error	Error en %
2T-2008	48.191	48.938	36.532	63.572	0.747	1,5%
3T-2008	52.606	52.516	39.282	70.451	-0.090	-0,2%
4T-2008	79.187	78.588	59.130	106.048	-0.599	-0,8%
1T-2009	79.164	81.791	59.114	106.014	2.627	3,2%

Cuadro E.5. Predicción dentro de la muestra actualizando a un paso.

Trimestre	Predicción	Valor Real	Int. Inf. 95	Int. Sup. 95	Error	Error en %
3T-2008	52.905	52.516	40.234	69.567	-0.389	-0,7%
4T-2008	79.123	78.588	59.297	105.577	-0.535	-0,7%
1T-2009	79.315	81.791	59.443	105.832	2.476	3,0%

Cuadro E.6. Predicción dentro de la muestra actualizando a un paso.

Trimestre	Predicción	Valor Real	Int. Inf. 95	Int. Sup. 95	Error	Error en %
4T-2008	78.884	78.588	60.173	103.414	-0.296	-0,4%
1T-2009	79.392	81.791	59.696	105.588	2.399	2,9%

Cuadro E.7. Predicción dentro de la muestra actualizando a un paso.

Trimestre	Predicción	Valor Real	Int. Inf. 95	Int. Sup. 95	Error	Error en %
1T-2009	79.449	81.791	60.789	103.838	2.342	2,9%

Cuadro E.8. Predicción dentro de la muestra actualizando a un paso.

APÉNDICE F: RESULTADOS DE TIPOLOGÍA DE TURISTAS

F.1. Ficha Técnica del MINTUR para Encuesta de Turismo Receptivo 2008

La investigación sobre Turismo Receptivo, a cargo de la División Investigación y Estadísticas del Ministerio de Turismo es una encuesta por muestreo de carácter continuo que se realiza prácticamente todos los días del año. Se realiza en forma ininterrumpida desde el año 1977.

La Encuesta de Turismo Receptivo es una investigación estadística por muestreo, dirigida a los pasajeros no residentes que visitan Uruguay. La Encuesta se realiza para conocer diversas características de los visitantes: país de origen, país de residencia, edad, sexo, composición del grupo con el que viajan (si viajan en grupo), motivos del viaje, lugar(es) de destino, tiempo de estadía y gasto per cápita, entre otras variables. Son elaboradas tomando en consideración las recomendaciones metodológicas de la Organización Mundial del Turismo así como también con las del Manual de Balanza de Pagos Revisión 5 del Fondo Monetario Internacional y del Sistema de Cuentas Nacionales 93 de Naciones Unidas, a fin de hacerlas compatibles con todo el espectro de estadísticas macroeconómicas.

Tipo de datos

Encuesta por muestreo.

Unidad de análisis

Grupo de Viaje.

Es una persona que viaja sola o un grupo de personas que comparten el destino del viaje (Uruguay), la época del viaje y los gastos del viaje (o la mayor parte de los gastos del viaje). Frecuentemente los integrantes del grupo viajan juntos (a la salida y/o a la entrada al país), pero la clave para definir el grupo es el gasto compartido.

Se realiza una entrevista personal con un informante calificado del grupo de viaje. Se considera informante calificado quien puede proporcionar datos sobre el gasto (generalmente el jefe del grupo u otra persona mayor de edad).

Cobertura

Cobertura geográfica

La investigación se realiza en todos los puestos de salida de viajeros internacionales (puestos de frontera y aeropuertos). Ellos son Colonia, Puerto, Carrasco, Chuy, Carmelo, Punta del Este, Fray Bentos, Paysandú, Otros, Salto, Aeropuerto Colonia, Piriápolis, Rivera, Nueva Palmira, Artigas, Bella Unión, Melo – Aceguá y Río Branco.

Universo

Visitantes (no residentes) que ingresan al país por cualquier motivo excepto la inmigración y el trabajo remunerado en Uruguay.

Muestreo

Procedimiento de muestreo

El diseño es:

- aleatorio
- estratificado
- en una sola etapa de selección
- de conglomerados

La estratificación es por puesto de salida. Los conglomerados son los grupos de viaje. Cada unidad de la muestra es el grupo de viaje.

Tamaño: 12.000 encuestas al año aproximadamente.

Distribución temporal de la muestra: En función del movimiento de pasajeros en los puestos de frontera. Por este motivo la muestra es más grande en los meses del verano y en Semana Santa.

Ponderaciones

Las variables que son utilizadas como ponderadores son **Coef** y **Coef_{tot}**, que se calculan trimestralmente en base a los datos proporcionados por la Dirección Nacional de Migraciones.

El expansor **Coef_{tot}** se utiliza para expandir las características generales de los visitantes, ya que cada registro es un grupo de personas. Este coeficiente incluye esta característica.

El expansor **Coef** se utiliza para expandir todo lo que tiene que ver con el gasto. Este es respondido para el total del grupo, por lo que el número de integrantes ya esta considerado en la respuesta.

Datos perdidos

Las variables que tienen valor 0 se consideran como sin datos.

En los casos de las variables de gasto, el valor sin datos, viene dado por el valor que asume la moneda de cada rubro de gasto.

Estos valores pueden ser: 1 - dólares americanos, 8 - sin gasto, 0 - sin datos.

En el gasto, en caso de que la moneda tome valor 0 - sin datos, este dato se ingresa.

Para la entrada de estos datos se utiliza la media del gasto por persona por destino por día de aquellos que han proporcionado datos, para cada rubro del gasto, salvo en el caso de transporte, en el cual también se toma en cuenta el tipo de transporte utilizado.

Este valor, multiplicado por la cantidad de personas del grupo y por la estadía es lo que se le asigna a cada valor de sin datos (0) en los gastos.

Notas

En el caso de gasto por alojamiento, para aquellos que concurren a Punta del Este a casa propia, se ingresa un gasto que surge de la media del gasto por persona por día en casa arrendada en el mismo destino. Esto fue a pedido del Banco Central del Uruguay.

F.2. Matriz de Correlación para el análisis de datos sin agregar

	Gen.Coef	Com.Coef	Vis.Coef	Gtot.Coef	Rec.Coef	Est.Coef
Gen.Coef	1	0,57	0,40	0,40	0,24	0,29
Com.Coef	0,57	1	0,11	0,73	0,28	0,80
Vis.Coef	0,40	0,11	1	0,11	0,10	0,15
Gtot.Coef	0,40	0,73	0,11	1	0,35	0,66
Rec.Coef	0,24	0,28	0,10	0,35	1	0,29
Est.Coef	0,29	0,80	0,15	0,66	0,29	1

Cuadro F.1. Matriz de Correlación.

F.3. Matriz de Saturación para el análisis de datos sin agregar

	Comp.1	Comp.2	Comp.3
Gente	-0,35	0,29	0,80
Comparte	-0,54	0,19	-0,00
Gtotal	-0,51	-0,00	-0,16
Recportap	-0,25	-0,93	0,22
Estadia	-0,49	0,05	-0,52

Cuadro F.2. Matriz de Saturación.

F.4. Matriz de Correlación para el análisis de datos agregados

	Estadia.prom.pg	Men15.rat	X1529.rat	X3064.rat	Mas.65rat	Destino1	Destino2
Estadia.prom.pg	1	-0,10	0,20	-0,12	0,07	0,33	-0,05
Men15.rat	-0,10	1	0,26	-0,68	0,17	-0,07	-0,12
X1529.rat	0,20	0,26	1	-0,78	-0,04	0,00	0,11
X3064.rat	-0,12	-0,68	-0,78	1	-0,41	0,04	-0,07
Mas.65rat	0,07	0,17	-0,04	-0,41	1	-0,04	0,13
Destino1	0,33	-0,07	0,00	0,04	-0,04	1	-0,26
Destino2	-0,05	-0,12	0,11	-0,07	0,13	-0,26	1
Destino3	0,27	-0,27	-0,04	0,12	0,07	-0,12	-0,17
Destino4	-0,30	0,02	-0,35	0,32	-0,23	-0,35	-0,20
Destino6	0,43	-0,03	0,24	-0,13	-0,04	0,32	-0,12
Destino7	-0,29	0,38	0,24	-0,37	0,11	-0,32	-0,16
Galoj.rat	0,09	-0,07	0,37	-0,13	-0,23	0,32	-0,33
Galim.rat	-0,26	0,06	-0,03	-0,05	0,13	-0,59	0,15
Gotr.rat	0,06	0,02	-0,34	0,16	0,15	0,03	0,23
	Destino3	Destino4	Destino6	Destino7	Galoj.rat	Galim.rat	Gotr.rat
Estadia.prom.pg	0,27	-0,30	0,43	-0,29	0,09	-0,26	0,06
Men15.rat	-0,27	0,02	-0,03	0,38	-0,07	0,06	0,02
X1529.rat	-0,04	-0,35	0,24	0,24	0,37	-0,03	-0,34
X3064.rat	0,12	0,32	-0,13	-0,37	-0,13	-0,05	0,16
Mas.65rat	0,07	-0,23	-0,04	0,11	-0,23	0,13	0,15
Destino1	-0,12	-0,35	0,32	-0,32	0,32	-0,59	0,03
Destino2	-0,17	-0,20	-0,12	-0,16	-0,33	0,15	0,23
Destino3	1	-0,28	0,16	-0,36	-0,28	-0,25	0,42
Destino4	-0,28	1	-0,27	-0,18	-0,25	0,35	0,04
Destino6	0,16	-0,27	1	-0,26	0,17	-0,18	-0,06
Destino7	-0,36	-0,18	-0,26	1	0,43	0,33	-0,61
Galoj.rat	-0,28	-0,25	0,17	0,43	1	-0,25	-0,82
Galim.rat	-0,25	0,35	-0,18	0,33	-0,25	1	-0,33
Gotr.rat	0,42	0,04	-0,06	-0,61	-0,82	-0,33	1

Cuadro F.3. Matriz de Correlación.

F.5. Matriz de Saturación para el análisis de datos agregados

	Comp.1	Comp.2	Comp.3	Comp.4	Comp.5
Estadia.prom.pg	-0,41	0,55	-0,17	-0,07	-0,10
Men15.rat	0,42	0,01	-0,52	0,63	0,10
X1529.rat	0,32	0,46	-0,37	-0,15	0,00
X3064.rat	-0,41	-0,31	0,74	-0,17	-0,01
Mas.65rat	0,01	-0,04	-0,69	0,00	-0,07
Destino1	-0,31	0,67	0,17	0,26	0,45
Destino2	-0,10	-0,33	-0,40	-0,63	0,51
Destino3	-0,62	0,04	-0,15	-0,08	-0,70
Destino4	0,10	-0,61	0,46	0,39	0,00
Destino6	-0,28	0,60	-0,05	0,00	-0,07
Destino7	0,87	0,07	-0,14	-0,06	-0,21
Galoj.rat	0,50	0,70	0,37	-0,15	-0,00
Galim.rat	0,47	-0,50	-0,06	-0,21	-0,19
Gotr.rat	-0,76	-0,39	-0,32	0,27	0,11

Cuadro F.4. Matriz de Saturación.

F.6. Historia de aglomeración según el algoritmo de Ward para el análisis de datos agregados

	Historia	Frec.	Rcuad	PsF	PsT
38	22 -19	5	0,894	21,45	7,70
39	31 14	8	0,884	21,37	5,90
40	32 -54	5	0,873	21,25	3,78
41	-13 12	3	0,861	21,21	11,60
42	36 -25	8	0,846	21,13	5,63
43	29 37	4	0,831	21,18	2,35
44	35 23	11	0,812	21,18	7,95
45	44 19	14	0,787	20,84	7,06
46	39 40	13	0,758	20,65	8,02
47	45 34	18	0,726	20,82	6,49
48	46 38	18	0,674	19,92	10,11
49	48 41	21	0,597	18,19	9,97
50	49 33	24	0,486	15,77	10,52
51	50 43	28	0,367	14,82	8,25
52	47 51	46	0,197	12,78	12,43
53	52 42	54	0	NaN	12,78

Cuadro F.5. Historia de aglomeración.

APÉNDICE G: CÓDIGOS R

G.1. Series de Tiempo

Como ejemplo se incluye el script utilizado en el R 2.9.2 para los datos trimestrales 1993.I-2009.I.

```
#####  
#carga package forecast  
library(forecast)  
#Cargo datos  
datos<-read.table('datproy.txt')  
dim(datos)  
#Expresamos los datos en miles  
enmiles=datos/1000  
#convierto los datos en una serie de tiempo, en miles  
serie<-ts(enmiles,start=1993,frequency=4)  
class(serie)  
plot(serie, xlab='Tiempo', main='Serie')  
#Análisis descriptivo de los datos  
summary(serie)  
var(serie)  
#Gráficos conjuntos Serie-Diagrama caja  
par(mfrow=c(2,1))  
plot(serie, xlab='Tiempo', main='Serie')  
boxplot(serie, main='Diagrama de caja')  
#####  
#1) IDENTIFICACIÓN  
#despliega valores de las autocorrelaciones (en lugar del plot)  
a<-acf(serie, lag.max =36, plot=T)  
#Despliega los gráficos: FAC y FACP  
par(mfrow=c(2,1))  
acf(serie, lag.max =36, main='Función Autocorrelación')  
pacf(serie, lag.max =36, main='Función Autocorrelación Parcial')  
#hace el gráfico y los correlogramas juntos  
tsdisplay(serie, lag.max= 36)  
#####  
#2) AJUSTE DEL MODELO  
#transformo la serie aplicando log y después le hago una dif. estacional, D=1.  
serie1<-transform(serie, logserie = log(serie))  
serie2=serie1[,2]  
serie3<-ts(serie2,start=1993,frequency=4)  
plot(serie3)  
#Análisis descriptivo de la serie transformada  
summary(serie3)  
var(serie3)  
#despliega valores de las autocorrelaciones de la serie transformada (en lugar del plot)  
acf.log<-acf(serie3, lag.max =24, plot=T)  
length(acf.log$lag)  
#Despliega los gráficos: FAC y FACP de la serie transformada  
par(mfrow=c(2,1))  
acf(serie3, lag.max =24, main='Función Autocorrelación Serie Transformada')  
pacf(serie3, lag.max =24, main='Función Autocorrelación Parcial Serie Transformada')  
#hace el gráfico y los correlogramas juntos de la serie transformada  
tsdisplay(serie3, lag.max= 36)  
#diferencio la serie transformada una vez en la parte estacional, D=1  
ds4serie<-diff(serie3, lag=4, differences = 1)  
#Análisis descriptivo de la serie diferenciada  
summary(ds4serie)  
var(ds4serie)  
#despliega valores de las autocorrelaciones de la serie diferenciada, D=1 (en lugar del plot)  
acf.ds<-acf(ds4serie, lag.max =24, plot=FALSE)  
length(acf.ds$lag)  
#Despliega los gráficos: FAC y FACP de la serie diferenciada, D=1  
par(mfrow=c(2,1))  
acf(ds4serie, lag.max =24, main='Función Autocorrelación Serie Transformada Dif D=1')  
pacf(ds4serie, lag.max =24, main='Función Autocorrelación Parcial Serie Transformada Dif D=1')  
#hace el gráfico y los correlogramas juntos de la serie diferenciada, D=1  
tsdisplay(ds4serie, lag.max= 36, main="D(log(serie))")
```

```
#####
#3) ESTIMACIÓN MV
#Según sugerencia de Demetra, empezamos a probar con un ARIMA (1,0,0)(0,1,1)(4) con intervenciones: Turismo,
tc2002, ls2002, tc2006 y ls2006.
#Significación de los coeficientes. Test t-student
reg1=read.table("xreg11.txt",header=TRUE,dec=",")
reg1b=ts(reg1,start=c(1993,1),frequency=4)
modelo1<-arima(serie, order = c(1, 0, 0), seasonal = list(order = c(0, 1, 1), period = 4), xreg=reg1b[1:65,], include.mean
= TRUE,method = "ML")
modelo1
attributes(modelo1)
t1=modelo1$coef/sqrt(diag(modelo1$var.coef))
t1
modelo1b<-arima(serie, order = c(1, 0, 0), seasonal = list(order = c(1, 1, 0), period = 4), xreg=reg1b[1:65,],
include.mean = TRUE,method = "ML")
modelo1b
attributes(modelo1b)
t1b=modelo1b$coef/sqrt(diag(modelo1b$var.coef))
t1b
modelo2<-arima(serie, order = c(1, 0, 0), seasonal = list(order = c(0, 1, 1), period = 4), xreg=reg1b[1:65,1:2],
include.mean = TRUE,method = "ML")
modelo2
attributes(modelo2)
t2=modelo2$coef/sqrt(diag(modelo2$var.coef))
t2
modelo2b<-arima(serie, order = c(1, 0, 0), seasonal = list(order = c(1, 1, 0), period = 4), xreg=reg1b[1:65,1:2],
include.mean = TRUE,method = "ML")
modelo2b
attributes(modelo2b)
t2b=modelo2b$coef/sqrt(diag(modelo2b$var.coef))
t2b
modelo3<-arima(log(serie), order = c(1, 0, 0), seasonal = list(order = c(0, 1, 1), period = 4), xreg=reg1b[1:65,],
include.mean = TRUE,method = "ML")
modelo3
attributes(modelo3)
t3=modelo3$coef/sqrt(diag(modelo3$var.coef))
t3
modelo4<-arima(log(serie), order = c(1, 0, 0), seasonal = list(order = c(0, 1, 1), period = 4), xreg=reg1b[1:65,1:2],
include.mean = TRUE,method = "ML")
modelo4
attributes(modelo4)
t4=modelo4$coef/sqrt(diag(modelo4$var.coef))
t4
modelo4b<-arima(log(serie), order = c(1, 0, 0), seasonal = list(order = c(1, 1, 0), period = 4), xreg=reg1b[1:65,1:2],
include.mean = TRUE,method = "ML")
modelo4b
attributes(modelo4b)
t4b=modelo4b$coef/sqrt(diag(modelo4b$var.coef))
t4b
reg2=read.table("xreg21.txt",header=TRUE,dec=",")
reg2b=ts(reg2,start=c(1993,1),frequency=4)
modelo5<-arima(serie, order = c(1, 0, 0), seasonal = list(order = c(0, 1, 1), period = 4), xreg=reg2b[1:65,], include.mean
= TRUE,method = "ML")
modelo5
t5=modelo5$coef/sqrt(diag(modelo5$var.coef))
t5
modelo6<-arima(serie, order = c(1, 0, 0), seasonal = list(order = c(0, 1, 1), period = 4), xreg=reg2b[1:65,1:2],
include.mean = TRUE,method = "ML")
modelo6
t6=modelo6$coef/sqrt(diag(modelo6$var.coef))
t6
modelo6b<-arima(serie, order = c(1, 0, 0), seasonal = list(order = c(1, 1, 0), period = 4), xreg=reg2b[1:65,1:2],
include.mean = TRUE,method = "ML")
modelo6b
t6b=modelo6b$coef/sqrt(diag(modelo6b$var.coef))
t6b
modelo6c<-arima(log(serie), order = c(1, 0, 0), seasonal = list(order = c(0, 1, 0), period = 4), xreg=reg2b[1:65,1:2],
include.mean = TRUE,method = "ML")
modelo6c
t6c=modelo6c$coef/sqrt(diag(modelo6c$var.coef))
t6c
modelo6c1<-arima(serie, order = c(1, 0, 0), seasonal = list(order = c(0, 1, 0), period = 4), xreg=reg2b[1:65,1:2],
include.mean = TRUE,method = "ML")
modelo6c1
t6c1=modelo6c1$coef/sqrt(diag(modelo6c1$var.coef))
t6c1
```

```

#no cumple con los supuestos de los residuos
modelo6d<-arima(log(serie), order = c(1, 0, 0), seasonal = list(order = c(0, 1, 0), period = 4), xreg=reg1b[1:65,1:2],
include.mean = TRUE,method ="ML")
modelo6d
t6d=modelo6d$coef/sqrt(diag(modelo6d$var.coef))
t6d
modelo6d1<-arima(serie, order = c(1, 0, 0), seasonal = list(order = c(0, 1, 0), period = 4), xreg=reg1b[1:65,1:2],
include.mean = TRUE,method ="ML")
modelo6d1
t6d1=modelo6d1$coef/sqrt(diag(modelo6d1$var.coef))
t6d1
#no cumple con los supuestos de los residuos
modelo6e<-arima(log(serie), order = c(1, 0, 0), seasonal = list(order = c(0, 1, 0), period = 4), xreg=reg1b[1:65,],
include.mean = TRUE,method ="ML")
modelo6e
t6e=modelo6e$coef/sqrt(diag(modelo6e$var.coef))
t6e
modelo7<-arima(log(serie), order = c(1, 0, 0), seasonal = list(order = c(0, 1, 1), period = 4), xreg=reg1[1:65,],
include.mean = TRUE,method ="ML")
modelo7
t7=modelo7$coef/sqrt(diag(modelo7$var.coef))
t7
modelo8<-arima(log(serie), order = c(1, 0, 0), seasonal = list(order = c(0, 1, 1), period = 4), xreg=reg1[1:65,1:2],
include.mean = TRUE,method ="ML")
modelo8
t8=modelo8$coef/sqrt(diag(modelo8$var.coef))
t8
#vemos que para los modelos 6c y 6d todos los coef. son significativos.
#####
#4.1) DIAGNÓSTICO para modelo 6d
#Residuos
res1<-modelo6d$residuals
sd(res1)
mean(res1)
var(res1)
par(mfrow=c(2,1))
plot(res1, main='Residuos')
plot(density(res1), main='Densidad de Residuos')
#Incorrelación
par(mfrow=c(2,1))
acf (res1, lag.max =36, main="Autocorrelograma Residuos")
pacf (res1, lag.max =36, main="Autocorrelograma Parcial Residuos")
#hace el gráfico y los correlogramas juntos de los residuos
tsdisplay(res1, lag.max= 36, main="Residuos")
#Testeo incorrelación de residuos
Box.test(res1,lag=4,type="Ljung-Box")
Box.test(res1,lag=8,type="Ljung-Box")
Box.test(res1,lag=12,type="Ljung-Box")
Box.test(res1,lag=16,type="Ljung-Box")
#Normalidad de los residuos
plot(density(res1))
qqnorm(res1) # qqplot
#test de normalidad de los residuos
sht<-shapiro.test(res1)
sht
#test de normalidad de jarque bera
jarque.bera.test(res1)
#Testeo homoscedasticidad de vza
res21<-res1*res1
par(mfrow=c(2,1))
acf (res21, lag.max =36, main="Autocorrelograma Residuos2")
pacf (res21, lag.max =36, main="Autocorrelograma Parcial Residuos2")
#4.2) DIAGNÓSTICO para modelo 6c
#Residuos
res2<-modelo6c$residuals
sd(res2)
mean(res2)
var(res2)
par(mfrow=c(2,1))
plot(res2, main='Residuos')
plot(density(res2), main='Densidad de Residuos')
#Incorrelación
par(mfrow=c(2,1))
acf (res2, lag.max =36, main="Autocorrelograma Residuos")
pacf (res2, lag.max =36, main="Autocorrelograma Parcial Residuos")
#hace el gráfico y los correlogramas juntos de los residuos

```

```

tsdisplay(res2, lag.max= 36, main="Residuos")
#Testeo incorrelación de residuos
Box.test(res2,lag=4,type="Ljung-Box")
Box.test(res2,lag=8,type="Ljung-Box")
Box.test(res2,lag=12,type="Ljung-Box")
Box.test(res2,lag=16,type="Ljung-Box")
#Normalidad de los residuos
plot(density(res2))
qqnorm(res2) # qqplot
#test de normalidad de los residuos
sht2<-shapiro.test(res2)
sht2
#test de normalidad de jarque bera
jarque.bera.test(res2)
#Testeo homoscedasticidad de vza
res22<-res2*res2
par(mfrow=c(2,1))
acf (res22, lag.max =36, main="Autocorrelograma Residuos2")
pacf (res22, lag.max =36, main="Autocorrelograma Parcial Residuos2")
#####
#5) PREDICCIÓN
#5.1.1. Predicción dentro de la muestra para el modelo 6d, reg1b
#Predicción a 4 pasos (h=4)
corte1<-window(log(serie), 1993, c(2008,1))
cortex1<-window(reg1b[1:61,1:2])
cortex1_p<-window(reg1b[62:65,1:2])
mod1<- arima(corte1, order = c(1, 0, 0), seasonal = list(order = c(0, 1, 0), period = 4),
  xreg = cortex1 , include.mean = TRUE, method = "ML")
mod1
t1=mod1$coef/sqrt(diag(mod1$var.coef))
t1
pred1<-forecast(mod1, h=4, conf=c(80,95),xreg=cortex1_p)
pred1
reales<-serie[c(62:65)]
reales
gof(pred1)
#Predicción a 1 paso actualizando
corte2<-window(log(serie), 1993, c(2008,1))
cortex2<-window(reg1b[1:61,1:2])
mod2<- arima(corte2, order = c(1, 0, 0), seasonal = list(order = c(0, 1, 0), period = 4),
  xreg =cortex2, include.mean = TRUE, method = "ML")
mod2
t2=mod2$coef/sqrt(diag(mod2$var.coef))
t2
corte2_p<-window(reg1b[62:65,1:2])
pred2<-forecast(mod2, h=4, conf=c(80,95),xreg=cortex2_p)
pred2
reales<-serie[c(62:65)]
reales
gof(pred2)
corte3<-window(log(serie), 1993, c(2008,2))
cortex3<-window(reg1b[1:62,1:2])
mod3<- arima(corte3, order = c(1, 0, 0), seasonal = list(order = c(0, 1, 0), period = 4),
  xreg =cortex3, include.mean = TRUE, method = "ML")
mod3
t3=mod3$coef/sqrt(diag(mod3$var.coef))
t3
corte3_p<-window(reg1b[63:65,1:2])
pred3<-forecast(mod3, h=3, conf=c(80,95),xreg=cortex3_p)
pred3
reales<-serie[c(63:65)]
reales
gof(pred3)
corte4<-window(log(serie), 1993, c(2008,3))
cortex4<-window(reg1b[1:63,1:2])
mod4<- arima(corte4, order = c(1, 0, 0), seasonal = list(order = c(0, 1, 0), period = 4),
  xreg =cortex4, include.mean = TRUE, method = "ML")
mod4
t4=mod4$coef/sqrt(diag(mod4$var.coef))
t4
corte4_p<-window(reg1b[64:65,1:2])
pred4<-forecast(mod4, h=2, conf=c(80,95),xreg=cortex4_p)
pred4
reales<-serie[c(64:65)]
reales
gof(pred4)

```

```

corte5<-window(log(serie), 1993, c(2008,4))
cortex5<-window(reg1b[1:64,1:2])
mod5<- arima(corte5, order = c(1, 0, 0), seasonal = list(order = c(0, 1, 0), period = 4),
  xreg = cortex5, include.mean = TRUE, method = "ML")
mod5
t5=mod5$coef/sqrt(diag(mod5$var.coef))
t5
cortex5_p<-window(reg1b[65:65,1:2])
pred5<-forecast(mod5, h=1, conf=c(80,95), xreg=cortex5_p)
pred5
reales<-serie[c(65:65)]
reales
gof(pred5)
#5.2.1. Predicción fuera de la muestra para el modelo 6d
corte6<-window(log(serie), 1993, c(2009,1))
cortex6<-window(reg1b[1:65,1:2])
mod6<- arima(corte6, order = c(1, 0, 0), seasonal = list(order = c(0, 1, 0), period = 4),
  xreg = cortex6, include.mean = TRUE, method = "ML")
mod6
t6=mod6$coef/sqrt(diag(mod6$var.coef))
t6
cortex6_p<-window(reg1b[66:69,1:2])
pred6<-forecast(mod6, h=4, conf=c(80,95), xreg=cortex6_p)
pred6
names(pred6)
plot(pred6)
gof(pred6) #calcula las medidas de ajuste para toda la muestra
#5.1.2. Predicción dentro de la muestra para el modelo 6c, reg2b
#Predicción a 4 pasos (h=4)
corte1<-window(log(serie), 1993, c(2008,1))
cortexa1<-window(reg2b[1:61,1:2])
cortexa1_p<-window(reg2b[62:65,1:2])
moda1<- arima(corte1, order = c(1, 0, 0), seasonal = list(order = c(0, 1, 0), period = 4),
  xreg = cortexa1 , include.mean = TRUE, method = "ML")
moda1
ta1=moda1$coef/sqrt(diag(moda1$var.coef))
ta1
preda1<-forecast(moda1, h=4, conf=c(80,95),xreg=cortexa1_p)
preda1
reales<-serie[c(62:65)]
reales
gof(preda1)
#Predicción a 1 paso actualizando
corte2<-window(log(serie), 1993, c(2008,1))
cortexb2<-window(reg2b[1:61,1:2])
modb2<- arima(corte2, order = c(1, 0, 0), seasonal = list(order = c(0, 1, 0), period = 4),
  xreg =cortexb2, include.mean = TRUE, method = "ML")
modb2
tb2=modb2$coef/sqrt(diag(modb2$var.coef))
tb2
cortexb2_p<-window(reg2b[62:65,1:2])
predb2<-forecast(modb2, h=4, conf=c(80,95),xreg=cortexb2_p)
predb2
reales<-serie[c(62:65)]
reales
gof(predb2)
corte3<-window(log(serie), 1993, c(2008,2))
cortexc3<-window(reg2b[1:62,1:2])
modc3<- arima(corte3, order = c(1, 0, 0), seasonal = list(order = c(0, 1, 0), period = 4),
  xreg =cortexc3, include.mean = TRUE, method = "ML")
modc3
tc3=modc3$coef/sqrt(diag(modc3$var.coef))
tc3
cortexc3_p<-window(reg2b[63:65,1:2])
predc3<-forecast(modc3, h=3, conf=c(80,95),xreg=cortexc3_p)
predc3
reales<-serie[c(63:65)]
reales
gof(predc3)
corte4<-window(log(serie), 1993, c(2008,3))
cortexd4<-window(reg2b[1:63,1:2])
modd4<- arima(corte4, order = c(1, 0, 0), seasonal = list(order = c(0, 1, 0), period = 4),
  xreg =cortexd4, include.mean = TRUE, method = "ML")
modd4
td4=modd4$coef/sqrt(diag(modd4$var.coef))
td4

```

```

cortexd4_p<-window(reg2b[64:65,1:2])
predd4<-forecast(modd4, h=2, conf=c(80,95),xreg=cortexd4_p)
predd4
reales<-serie[c(64:65)]
reales
gof(predd4)
corte5<-window(log(serie), 1993, c(2008,4))
cortexe5<-window(reg2b[1:64,1:2])
mode5<- arima(corte5, order = c(1, 0, 0), seasonal = list(order = c(0, 1, 0), period = 4),
xreg = cortexe5, include.mean = TRUE, method = "ML")
mode5
te5=mode5$coef/sqrt(diag(mode5$var.coef))
te5
cortexe5_p<-window(reg2b[65:65,1:2])
prede5<-forecast(mode5, h=1, conf=c(80,95), xreg=cortexe5_p)
prede5
reales<-serie[c(65:65)]
reales
gof(prede5)
#5.2.2. Predicción fuera de la muestra para el modelo 6c
corte6<-window(log(serie), 1993, c(2009,1))
cortexf6<-window(reg2b[1:65,1:2])
modf6<- arima(corte6, order = c(1, 0, 0), seasonal = list(order = c(0, 1, 0), period = 4),
xreg = cortexf6, include.mean = TRUE, method = "ML")
modf6
tf6=modf6$coef/sqrt(diag(modf6$var.coef))
tf6
cortexf6_p<-window(reg2b[66:69,1:2])
predf6<-forecast(modf6, h=4, conf=c(80,95), xreg=cortexf6_p)
predf6
names(predf6)
plot(predf6)
gof(predf6) #calcula las medidas de ajuste para toda la muestra
#####

```

G.2. Tipología de Turistas

Se incluye el script utilizado en el R 2.9.2 para el análisis de los datos sin agregar.

```

#####
###ACP
acppr=read.table('acp primera parte.txt',sep='\t',dec=',', header=T)
dim(acppr)
names(acppr)
attach(acppr)
str(acppr)
acppr$trimestre=as.factor(acppr$trimestre)
#veo que son todas variables cuantitativas menos trimestre:
class(gente)
class(comparte)
class(visito)
class(gtotal)
class(recportap)
class(estadia)
class(coef)
class(trimestre)
#análisis descriptivo de datos
summary(acppr)
#saco la variable visita, coeftot y trimestre
acppr=acppr[,-1]
acppr=acppr[,-3]

```

```

acppr=acppr[,-7]
str(acppr)
#####
ACP
library(ade4)
library(ade4TkGUI)
ade4TkGUI()
source('acp.R')
source('standard.R')
#estandarizo datos
acpprst<-standard(acppr)
#acpprst=cbind(acppr[,1],acpprst)
head(acpprst)
summary(acpprst)
str(acpprst)
#ACP con dudi
a=acppr$coef
acpprst=acpprst[,1:5]
head(acpprst)
tur=dudi.pca(acpprst, row.w = a, col.w = rep(1, ncol(acpprst)), center = FALSE, scale = FALSE, scannf = TRUE, nf = 5)
str(tur)
#ineria
tur[4]
ine=unlist(tur[4])
deno=sum(unlist(tur[4]))
ineria=ine/deno
inep=cumsum(unlist(ineria))
inep
#me quedo con 2 o 3 componentes: 78,8% o 91,3%
sat=tur$c1 #matriz de saturación
sat
comp=tur$li #comp:coordenadas de los individuos en los componentes
head(comp)
dim(comp)
#####
#CLUSTERS
library(cluster)
source('indicadores.R')
library(vegan)
#uso comp
#head(comp)
#dim(comp)
#####
#debido al alto numero de obs=10415 se usa clustering no jerarquico
#probando con clustering no jerarquico: kmeans
#la funcion cascadeKM es un wrapper de la función kmeans
#me ayuda a determinar el numero de clusters
cas1=cascadeKM(comp, 3, 6, iter = 10, criterion = "calinski")
str(cas1)
cas1$results
#elijo 3 o 5 grupos.

```

```

#cas1$size
plot(cas1)
#k=3
cl1=kmeans(comp, 3, iter.max = 10, nstart = 1, algorithm = "Hartigan-Wong")
str(cl1)
cl1$size
#k=5
cl2=kmeans(comp, 5, iter.max = 10, nstart = 1, algorithm = "Hartigan-Wong")
str(cl2)
cl2$size
#elegimos trabajar con 5 grupos.
#agrega a los 3 comps originales la columna que indica a que cluster va cada individuo
clus1<-cbind(comp,cl2$cluster)
head(clus1)
dim(clus1)
write.table(clus1,'clus1.txt',sep='\t',dec= '.')
clus1=read.table('clus1.txt',sep='\t',dec='.', header=T)
#hace summary de cada comp por grupo.
sum1<-by(clus1,clus1[,4],summary)
#diagramas de caja por grupos
par(mfrow=c(2,2))
boxplot(clus1[,1]~clus1[,4],main='Componente 1',col='lightgreen')
boxplot(clus1[,2]~clus1[,4],main='Componente 2',col='lightgreen')
boxplot(clus1[,3]~clus1[,4],main='Componente 3',col='lightgreen')
#####

```

Se incluye el script utilizado en el R 2.9.2 para el análisis de los datos agregados.

```

#####
library(foreign)
library(survey)
library(xtable)
datosacp=read.table('2p1t.txt',sep='\t',dec=',', header=T)
datosacp$lugsalid=as.factor(datosacp$lugsalid)
datosacp$destino=as.factor(datosacp$destino)
dim(datosacp)
names(datosacp)
attach(datosacp)
class(gente)
class(gtotal)
class(estadia)
class(lugsalid)
class(destino)
#stratified sample
dstrat1<-svydesign(id=~1, strata=~lugsalid, weights=~coef, data=datosacp)
summary(dstrat1)
promedio=svymean(estadia,dstrat1,def=TRUE)
total=svytotal(estadia,dstrat1,def=TRUE)
totalg=svytotal(gente,dstrat1,def=TRUE)
total1=svyby(~estadia,~lugsalid,dstrat1,svymean,def=TRUE)

```



```

total2=svyby(~gente,~lugsalid,dstrat1,svytotal,def=TRUE)
total3=svyby(~destino,~lugsalid,dstrat1,svymean,def=TRUE)
total3a=svyby(~destino,~lugsalid,dstrat1,svytotal,def=TRUE)
total4=svyby(~destino,~lugsalid,dstrat2,svytotal,def=TRUE)
dstrat2<-svydesign(id=~1, strata=~lugsalid, weights=~coefot, data=datosacp)
round(total4[,2:7],0)
#####
total5=svyby(~menos15,~lugsalid,dstrat1,svytotal,def=TRUE)
#round(total5[,2],0)
total6=svyby(~entre152,~lugsalid,dstrat1,svytotal,def=TRUE)
#round(total6[,2],0)
total7=svyby(~entre306,~lugsalid,dstrat1,svytotal,def=TRUE)
#round(total7[,2],0)
total8=svyby(~mas65,~lugsalid,dstrat1,svytotal,def=TRUE)
#round(total8[,2],0)
total9=svyby(~fmenos15,~lugsalid,dstrat1,svytotal,def=TRUE)
#round(total9[,2],0)
total10=svyby(~f15a29,~lugsalid,dstrat1,svytotal,def=TRUE)
#round(total10[,2],0)
total11=svyby(~f30a65,~lugsalid,dstrat1,svytotal,def=TRUE)
#round(total11[,2],0)
total12=svyby(~fmas64,~lugsalid,dstrat1,svytotal,def=TRUE)
#round(total12[,2],0)
#####
totalg1=svyby(~galoj,~lugsalid,dstrat1,svytotal,def=TRUE)
totalg2=svyby(~galim,~lugsalid,dstrat1,svytotal,def=TRUE)
totalg3=svyby(~gtrans,~lugsalid,dstrat1,svytotal,def=TRUE)
totalg4=svyby(~gcomp,~lugsalid,dstrat1,svytotal,def=TRUE)
totalg5=svyby(~gresto,~lugsalid,dstrat1,svytotal,def=TRUE)
nuevatabla=cbind(total1[,1:2],total5[,2],total6[,2],total7[,2],total8[,2],total9[,2],total10[,2],total11[,2],total12[,2],totalg1[,2],
totalg2[,2],totalg3[,2],totalg4[,2],totalg5[,2])
round(nuevatabla[,3:15])
tablafinal=cbind(nuevatabla[,1:2],round(nuevatabla[,3:15]))
write.table(tablafinal,'tablafinal.txt',sep='\t',dec=',')
#####
library(ade4)
library(ade4TkGUI)
source('acp.R')
source('standard.R')
datos14v=read.table('14vratios.txt',sep='\t',dec=',', header=T)
dim(datos14v)
datos14v<-as.data.frame(datos14v)
#estandarizo datos
datos14vst<-standard(datos14v)
summary(datos14vst)
#ACP COMÚN
b<-acp(datos14vst, Ncomps=8)
norma<-b[6]
norma
write.table(norma,'norma.txt',sep='\t',dec=',')
ineracia<-b[[1]]

```

```

inercia
matsat<-b[[4]]
comp<-b[[5]]
cosenos<-b[[7]]
cos1<-cosenos[cosenos[,1]<0.5,1] #cos2 en comp 1: hay
cos2<-cosenos[cosenos[,2]<0.5,2] #cos2 en comp 2: hay
cos3<-cosenos[cosenos[,3]<0.5,3] #cos2 en comp 3: hay
cos4<-cosenos[cosenos[,4]<0.5,4] #cos2 en comp 4: hay
cos5<-cosenos[cosenos[,5]<0.5,5] #cos2 en comp 5: hay
cos6<-cosenos[cosenos[,6]<0.5,6] #cos2 en comp 6: hay
cos7<-cosenos[cosenos[,7]<0.5,7] #cos2 en comp 7: hay
cos8<-cosenos[cosenos[,8]<0.5,8] #cos2 en comp 8: hay
#vemos que el individuo 52 tiene norma muy alta (93.314470) lo que indica que el individuo 52 (Artigas4T) es un valor
atípico.
#Por lo que corremos un nuevo acp sin el individuo 52.
#el individuo 52 es raro, entonces lo ponemos en Fsup
c<-acp(datos14vst, Fsup=52, Ncomps=8)
inercia1<-c[[1]]
inercia1
write.table(inercia1,'inercia1.txt',sep='\t',dec=',')
norma1<-c[6]
norma1
comp1<-c[[5]]
comp1
head(comp1)
dim(comp1)
write.table(comp1,'comp.txt',sep='\t',dec=',')
matcorr1<-c[[2]]
matcorr1
write.table(matcorr1,'matcorr.txt',sep='\t',dec=',')
#matriz de saturación
matsat1<-c[[4]]
aaa<-as.data.frame(matsat1)
dim(aaa)
sat<-aaa[1:14,15:22]##matriz de corr var-comp
sat
write.table(sat,'saturacion.txt',sep='\t',dec=',')
#Vemos cuantos individuos hay con cos2<0.5 en cada eje
cosenos1<-c[[7]]
write.table(cosenos1,'cosenos.txt',sep='\t',dec=',')
cos1<-cosenos1[cosenos1[,1]<0.5,1] #cos2 en comp 1: hay
cos2<-cosenos1[cosenos1[,2]<0.5,2] #cos2 en comp 2: hay
cos3<-cosenos1[cosenos1[,3]<0.5,3] #cos2 en comp 3: hay
cos4<-cosenos1[cosenos1[,4]<0.5,4] #cos2 en comp 4: hay
cos5<-cosenos1[cosenos1[,5]<0.5,5] #cos2 en comp 5: hay
cos6<-cosenos1[cosenos1[,6]<0.5,6] #cos2 en comp 6: hay
cos7<-cosenos1[cosenos1[,7]<0.5,7] #cos2 en comp 7: hay
cos8<-cosenos1[cosenos1[,8]<0.5,8] #cos2 en comp 8: hay
#####
#CLUSTERS
comp5=comp1[,-(6:8)]

```

```

library(cluster)
source('standard.R')
source('indicadores.R')
#Métodos de clasificación
#WARD
ward<-agnes(comp5, metric = "euclidean", stand = FALSE, method = "ward")
#indicadores
indw<-indicadores(ward[4],comp5,imprime=15)
write.table(indw,'indward.txt',sep='\t',dec='.')
#dendrograma
plot(ward,which=2, main='Algoritmo de Ward')
#VECINO MAS CERCANO
cercano<-agnes(comp5, metric = "euclidean", stand = FALSE, method = "single")
indc<-indicadores(cercano[4],comp5,imprime=10)
plot(cercano,which=2, main='Vecino más cercano')
#VECINO MAS LEJANO
lejano<-agnes(comp5, metric = "euclidean", stand = FALSE, method = "complete")
indl<-indicadores(lejano[4],comp5,imprime=10)
plot(lejano,which=2, main='Vecino más lejano')
#nos quedamos con 3 o 5 grupos con método ward o 2,5 o 6 grupos con método del vecino más lejano.
#Creo columna 'grupos'
#ward
k<-3
grupos<-as.factor(cutree(ward[4],k))
frec<-table(grupos)
frec
#ward
k<-5
grupos<-as.factor(cutree(ward[4],k))
frec<-table(grupos)
frec
#lejano
k<-2
grupos<-as.factor(cutree(lejano[4],k))
frec<-table(grupos)
frec
#lejano
k<-5
grupos<-as.factor(cutree(lejano[4],k))
frec<-table(grupos)
frec
#lejano
k<-6
grupos<-as.factor(cutree(lejano[4],k))
frec<-table(grupos)
frec
#se decidió elegir 3 grupos con ward ya que proporciona una estructura de grupos más pareja.
k<-3
grupos<-as.factor(cutree(ward[4],k))
frec<-table(grupos)
frec

```

```

#agrega a los datos originales la columna 'grupos'
clus<-cbind(comp5,grupos)
write.table(clus,'clus.txt',sep='\t',dec='.')
#veo los nombres de los puntos de salida q están en cada grupo
g1<-row.names(clus[clus[,6]==1,])
g2<-row.names(clus[clus[,6]==2,])
g3<-row.names(clus[clus[,6]==3,])
#hace summary de cada variable por grupo.
sum<-by(clus,clus[,6],summary)
#diagramas de caja por grupos
par(mfrow=c(3,2))
boxplot(clus[,1]~clus[,6],main='Componente 1',col='lightgreen')
boxplot(clus[,2]~clus[,6],main='Componente 2',col='lightgreen')
boxplot(clus[,3]~clus[,6],main='Componente 3',col='lightgreen')
boxplot(clus[,4]~clus[,6],main='Componente 4',col='lightgreen')
boxplot(clus[,5]~clus[,6],main='Componente 5',col='lightgreen')
#####
mat=read.table('mat.txt',sep='\t',dec=',', header=T)
mat1=cbind(mat,clus[,6])
write.table(mat1,'mat1.txt',sep='\t',dec='.')
#####

```