

Trabajo final

Matias Bajac - Aris Sarkisian

Metodologia

En ocasión del proyecto final del curso de Series Cronologicas, se propuso poner en práctica los conocimientos adquiridos durante el semestre y aplicarlos en una base de datos real. La cual constaba inicialmente con cantidad de turistas mensuales desde el año 1990 a marzo del 2024. Adicionalmente se creó 2 variables indicadoras mas (turismo y carnaval) la cual nos es de utilidad a la hora de estimar el modelo.

En primer instancia, se hara una exploración de datos para ver la evolución del turismo brasilero a lo largo del tiempo, como tambien se estudiara como una primera aproximación mediante los graficos de autocorrelacion si existe o no estacionariedad en la serie original, en cualquier caso, estos gráficos no resultan definitivos para determinar si el proceso es trend stationary o difference stationary Luego se procede a realizar una transformacion logaritmica a la serie debido a que la varianza aumenta con la media de la serie. El siguiente paso es realizar una diferencia estacional, para luego dar lugar a la intervención de outliers, en particular poner a suma atención en la pandemia del 2020, en el cual será intervenido manualmente. Como algo adicional, se hizo un test de raices unitarias, en el cual fue de utilidad ya que nos ayudo para agregar una constante y drift a los modelos estimados. En principio se presentan 2 modelos, uno solo con los datos atipicos, y el otro agregando las variables de carnaval y turismo. Luego se crean otros 2 modelos, esta vez analizando por un lado el efecto de la pandemia y los outliers, y finalmente agregando estos efectos a las variables carnaval y turismo. En todos los casos se hace el diagnostico de normalidad. Finalmente se utilizan estos 2 ultimos modelos mencionado para realizar predicciones a 12 pasos. Por ultimo se hace la validacion de las predicciones.

Serie elegida

Luego de contemplar distintas posibilidades, se determinó que la serie a utilizar sería la referida al turismo en Uruguay, específicamente el proveniente de Brasil. Esta es una serie mensual que contabiliza los visitantes ingresados a Uruguay, diferenciando según nacionalidad, que se encuentra disponible para el público por cortesía del Ministerio de Turismo. La serie inicia en enero del 1990 y tiene datos hasta marzo de 2024, lo que otorga un total de 411 observaciones para trabajar.

Para familiarizarse con la serie, se presentan unos pocos elementos que ayudarán a comprender el comportamiento que ha evidenciado el turismo receptivo durante este período.

En el primer gráfico, se presenta la evolución de los valores en el tiempo. Se puede ver que continuamente sube y baja, lo que se debe al comportamiento estacional del turismo, que hay una tendencia al alta en la cantidad total de turismo, en todas las estaciones, y sobre todo, resalta el efecto completamente negativo de la pandemia en el turismo, llegando a valores cercanos a 0, y tardando unos años en volver a los valores anteriores. En el segundo gráfico se estudia el comportamiento estacional del turismo. Considerando que el mayor atractivo turístico del Uruguay son las playas y los balnearios en general, tiene todo el sentido del mundo que en promedio los meses con mayor cantidad de turistas sean los asociados con el verano. El otro valor alto es julio, que coincide con las vacaciones de Julio, en la que mucha gente aprovecha para viajar, aún sin las mejores temperaturas.

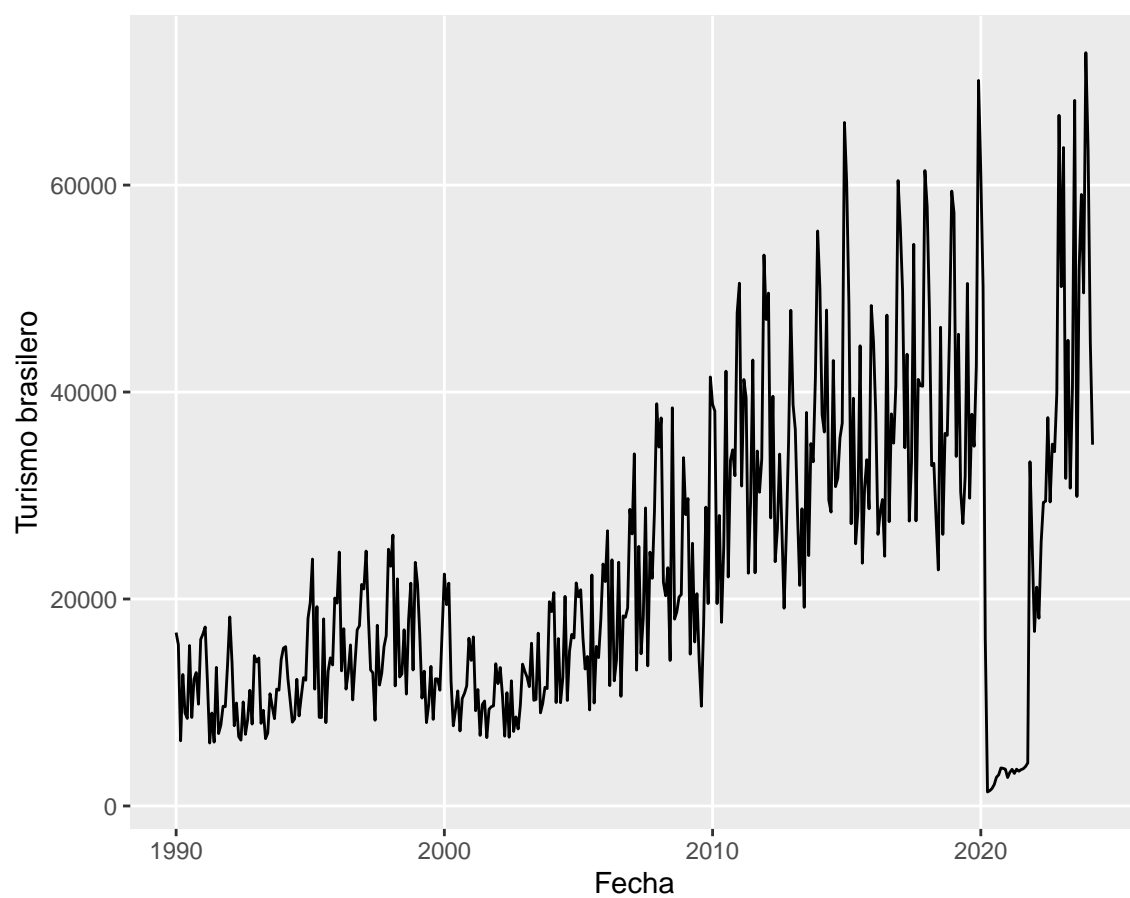


Figure 1: Evolución del turismo mensual proveniente de Brasil entre 1990 y 2024

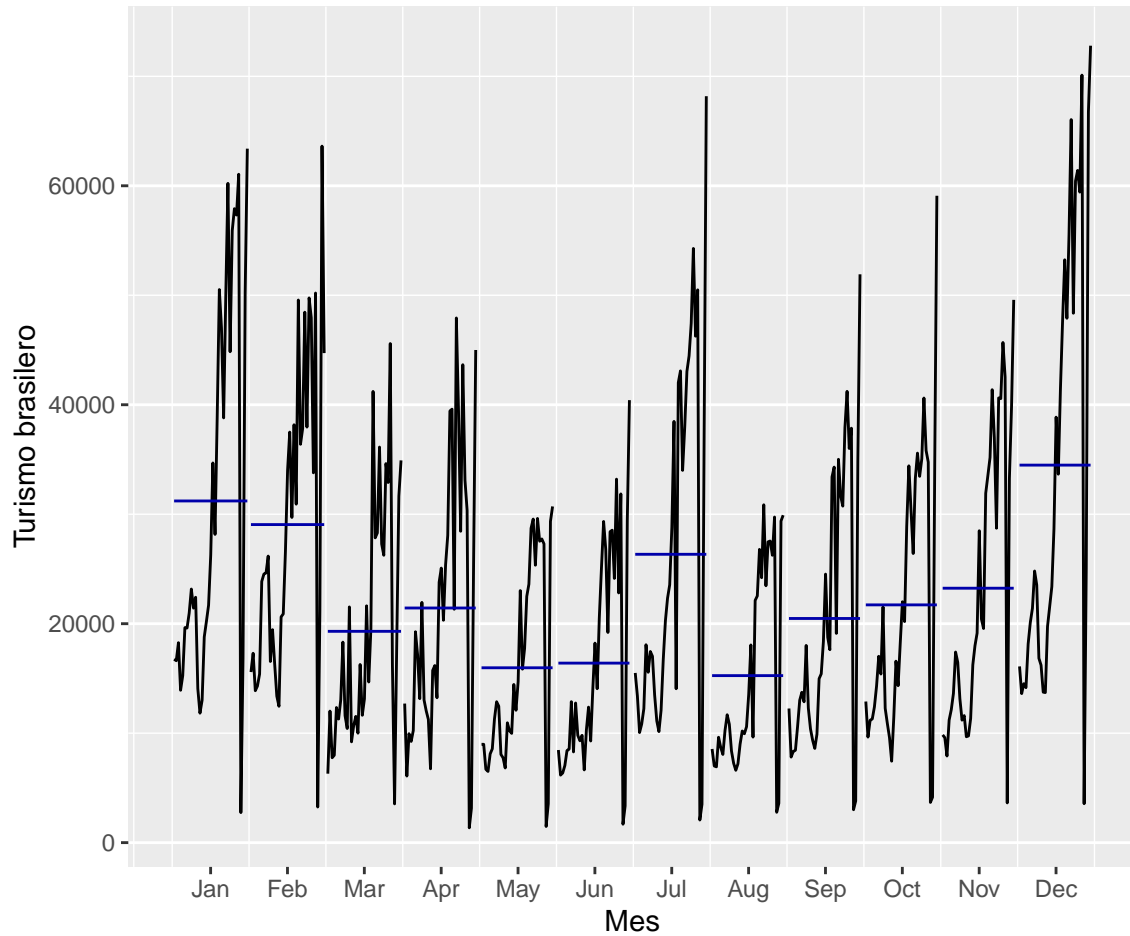


Figure 2: Comportamiento mensual del turismo proveniente de Brasil entre 1990 y 2024

Identificación del modelo

Para tener la capacidad de elegir el mejor modelo posible, es necesaria llevar a cabo una etapa de trabajar con la serie, hasta encontrar alguna operación que lleve a pensar que la serie pueda ser estacionaria.

sin hacer nada

En primer lugar se analiza la serie original, para estudiar si la serie es estacionaria o no. Y, al analizar, el FAC, se llega rápidamente a la conclusión de que la serie no es estacionaria, veiendo que los residuos están muy correlacionados entre si.

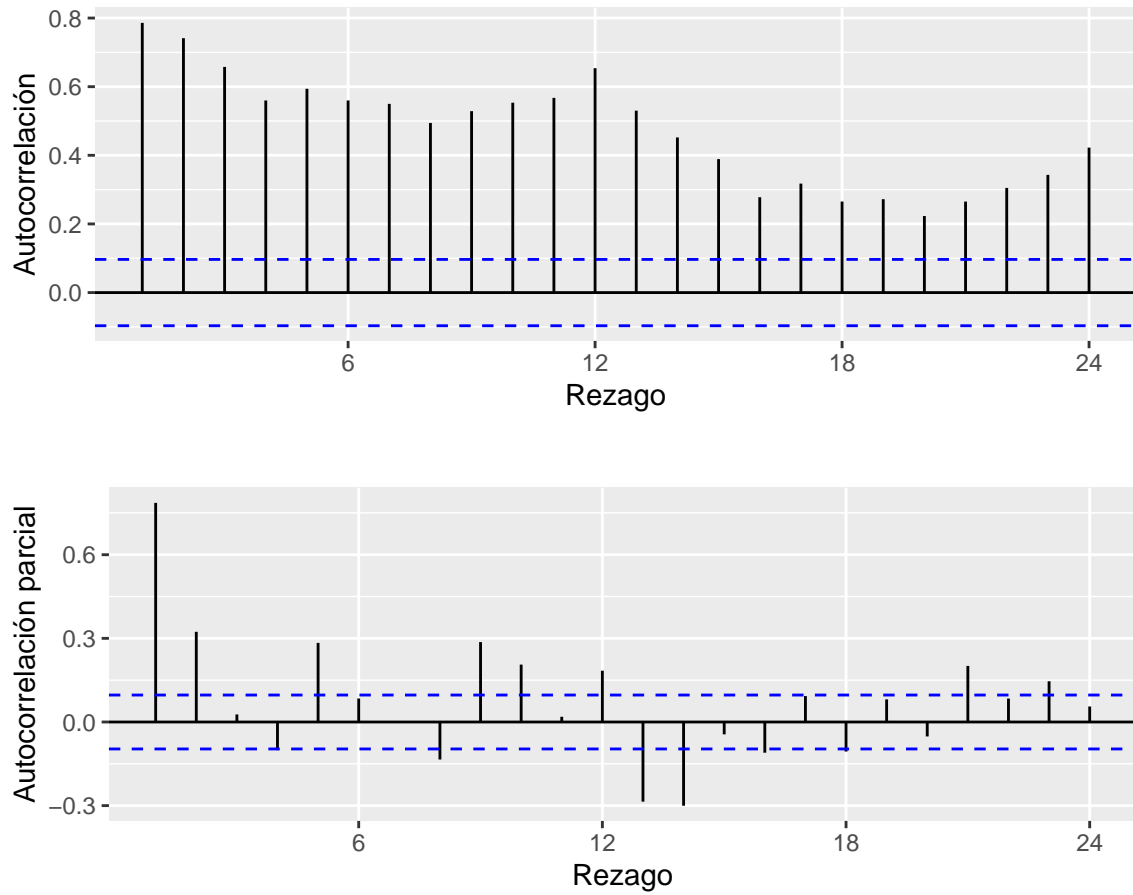


Figure 3: Funciones de Autocorrelacion y Autocorrelación parcial estimadas del turismo mensual proveniente de Brasil entre 1990 y 2024

Logaritmo

La serie original presenta una varianza ascensiediente en el tiempo, siendo que cada había más distancia entre los meses con mayor cantidad de turistas brasileiros y lo que contaban con menos. Entonces, se propone una transformación logarítmica, para ver si ayuda a estabilizar la varianza. Viendo la evolución de la serie, se podría decir que se logró reducir la varianza. De igual manera, analizando los FAC Y FACP, el proceso sigue sin ser estacionario.

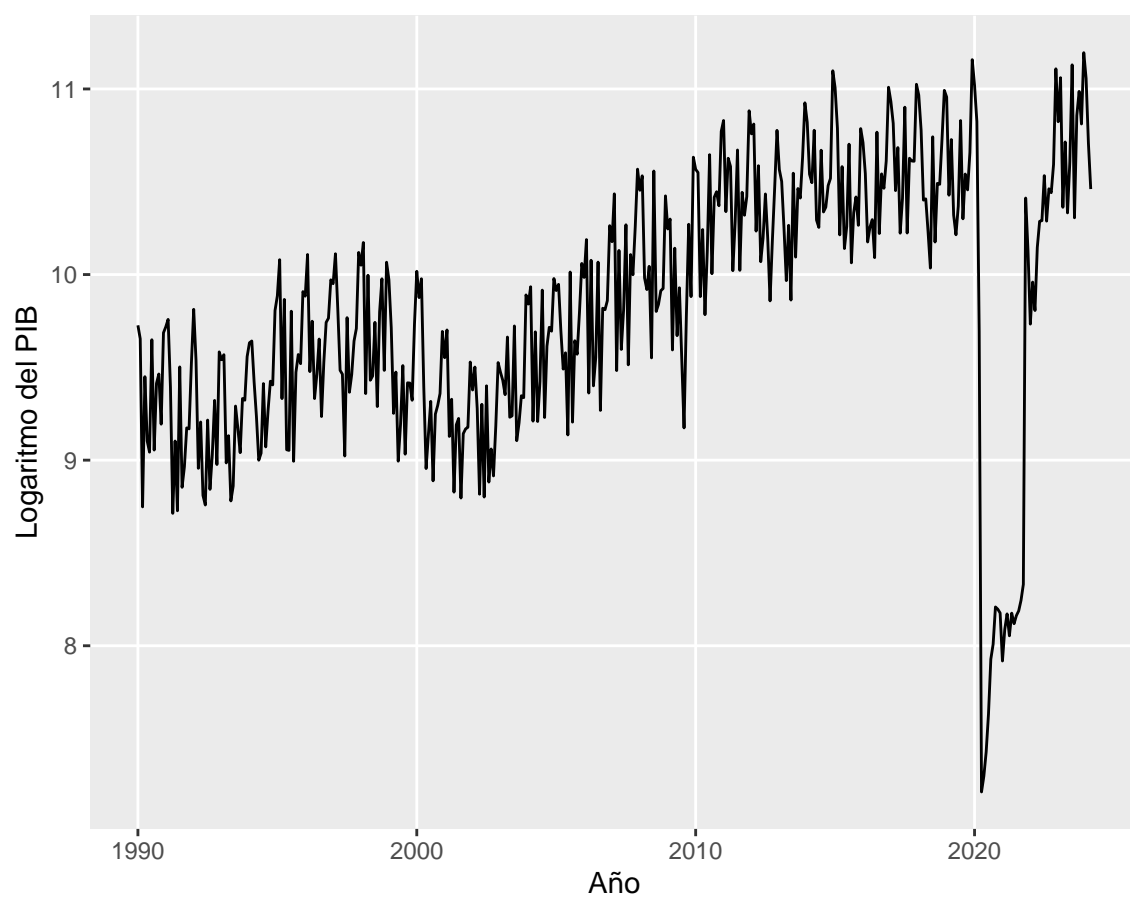


Figure 4: Evolución del logaritmo del turismo mensual proveniente de Brasil entre 1990 y 2024

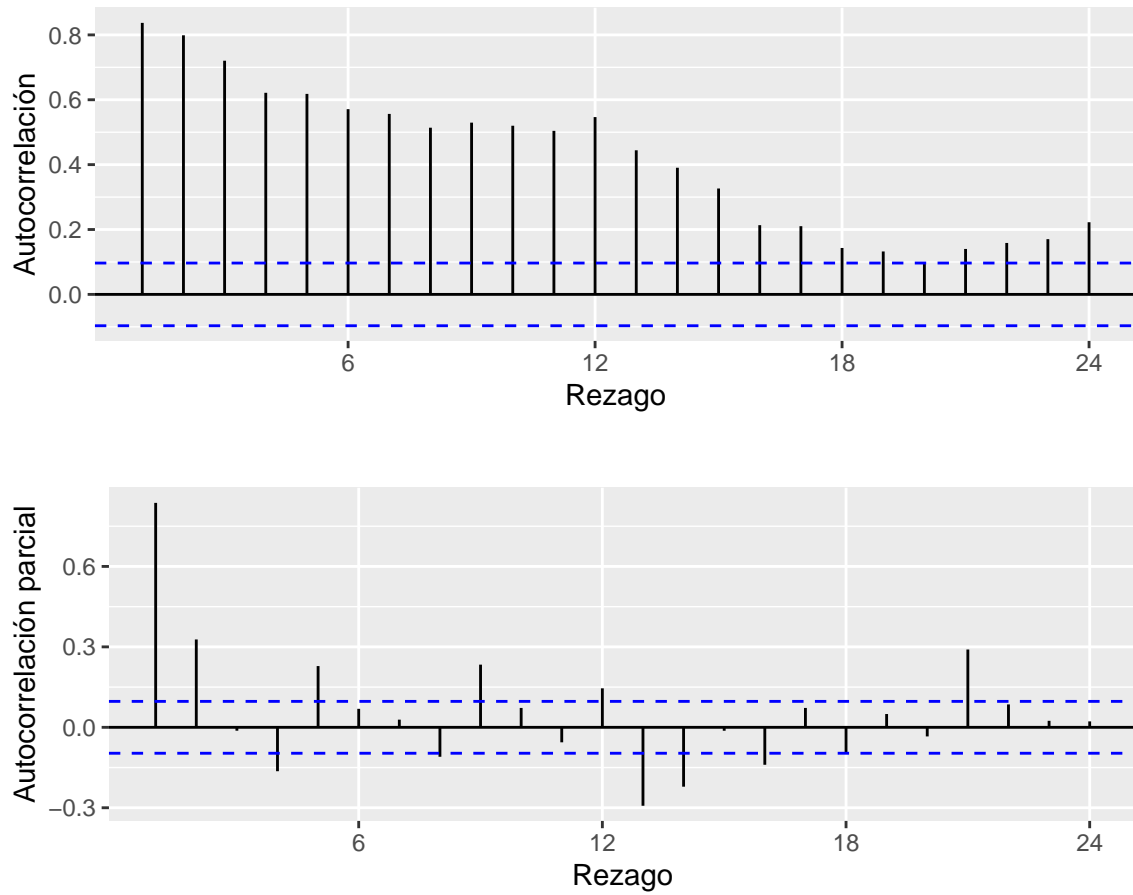


Figure 5: Funciones de Autocorrelacion y Autocorrelación parcial estimadas del logaritmo del turismo mensual proveniente de Brasil entre 1990 y 2024

Primera diferencia estacional

Ya había quedado establecido en la sección de exploración que el proceso contiene un fuerte componente estacional. Entonces, se procede a incorporar esta información al modelo propuesto. Pudiendo observar los gráficos propuestos, dejando de lado el efecto de la pandemia, la serie comienza a aparecerse a algo estacionario. Pasando a estudiar los resultados a los que se llega con los FAC Y FACP, la serie parece tener un predominio de la parte MA, indicando la parte estacionaria en la FACP, que desciende casi instantaneamente, y teniendo una FAC que converge asintoticamente a 0.

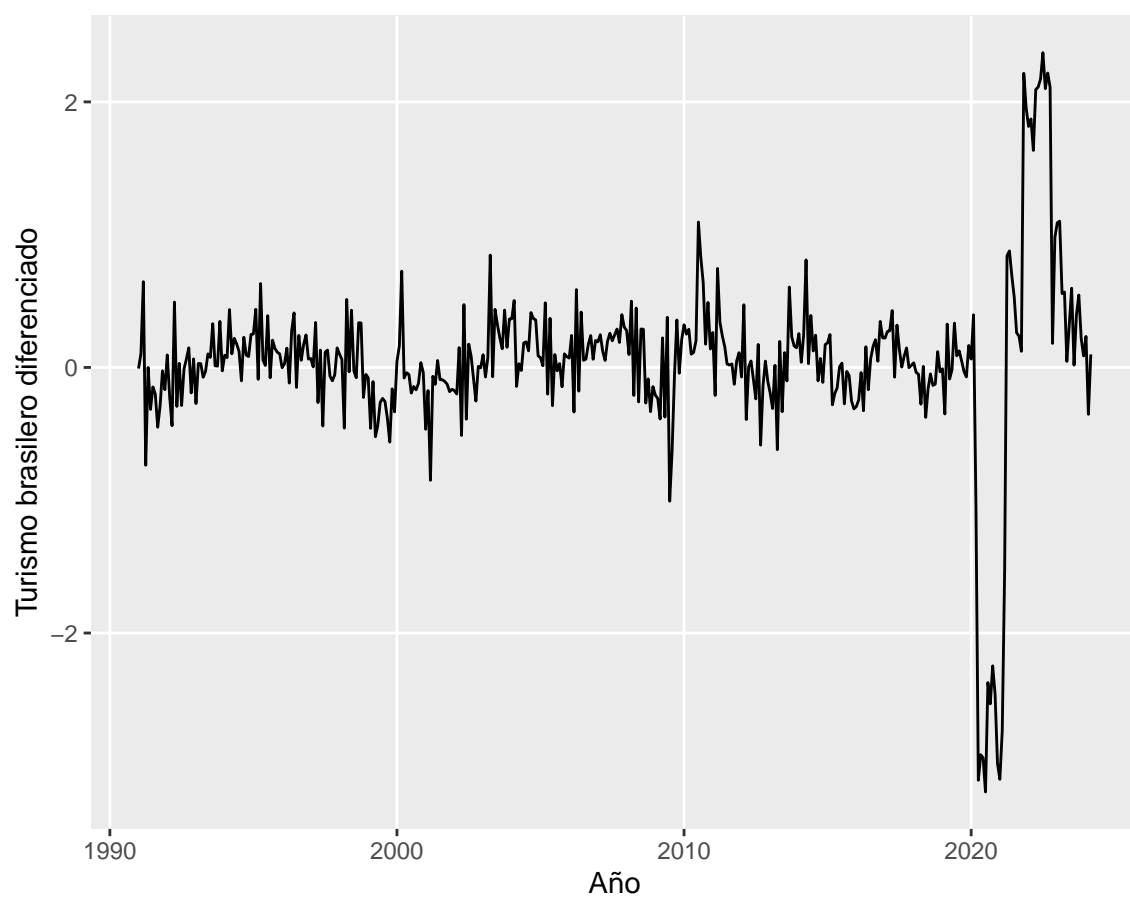


Figure 6: Evoución de la primera diferencia estacional del logaritmo del turismo mensual proveniente de Brasil entre 1990 y 2024

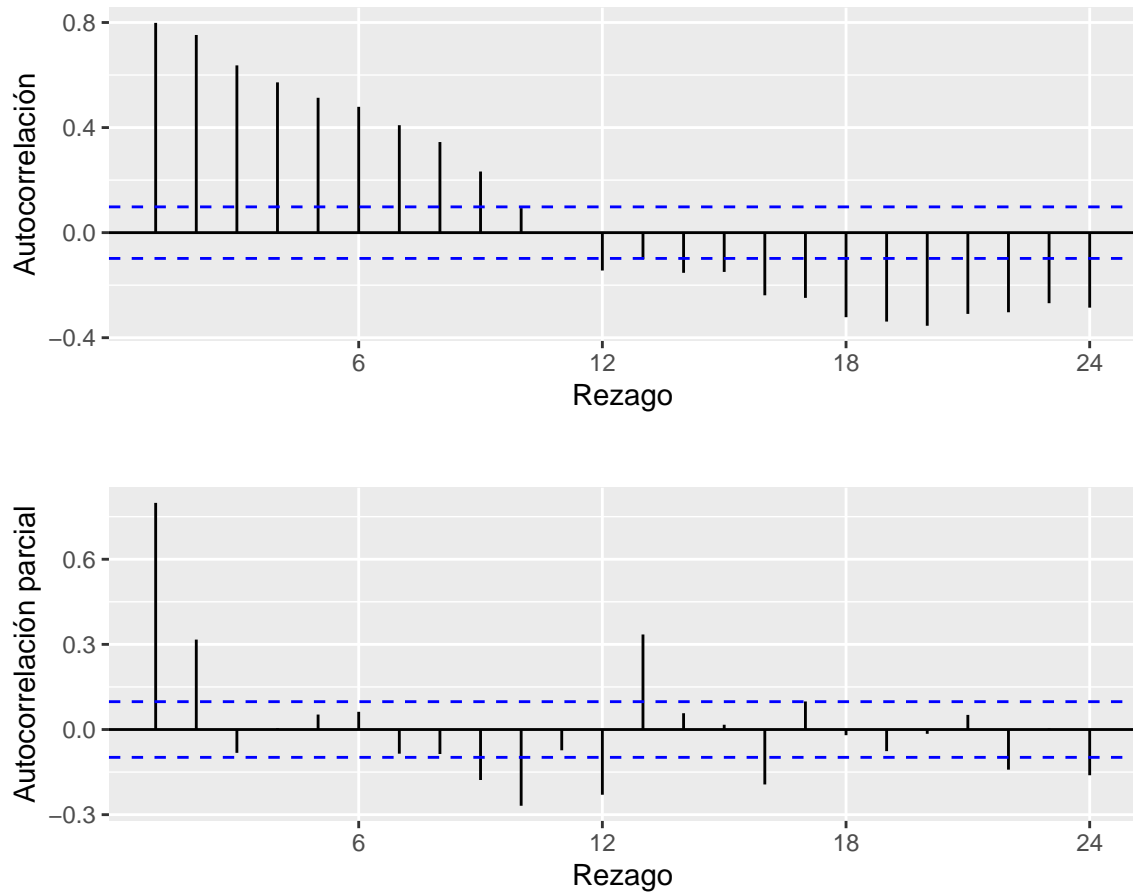


Figure 7: Funciones de Autocorrelacion y Autocorrelación parcial estimadas de la primera diferencia estacional del logaritmo del turismo mensual proveniente de Brasil entre 1990 y 2024

Detección outliers

Entendiendo que la detección de outliers generalmente es un paso que se presenta más adelante en el proceso de construcción del modelo, la presencia de un valor tan fuertemente atípico como la pandemia lleva a pensar que en realidad ningún modelo va a funcionar de gran manera si no se le avisa de antemano de la necesaria intervención de este conjunto de valores. El modelo elegido para encontrar esos outliers es meramente provisional con el único objetivo de detectar estos outliers, bajo el supuesto de que cualquier modelo razonablemente bueno logrará capturar el efecto de la pandemia. En este caso, además detectó otros outliers que no tienen nada que ver con la pandemia. Llegado el momento de realizar el modelo final, en caso de que no sean significativas estas intervenciones, simplemente se quitan.

##	type	ind	time	coefhat	tstat
## 1	A0	16	1991:04	-0.6017040	-4.033148
## 2	A0	123	2000:03	0.6705944	4.612682
## 3	TC	235	2009:07	-0.8418000	-5.571966
## 4	LS	364	2020:04	-2.8207883	-18.250996
## 5	LS	368	2020:08	0.6028997	3.901951
## 6	TC	372	2020:12	-0.6625694	-4.293638
## 7	LS	383	2021:11	2.1533099	12.400372
## 8	TC	384	2021:12	-0.9834705	-5.676455

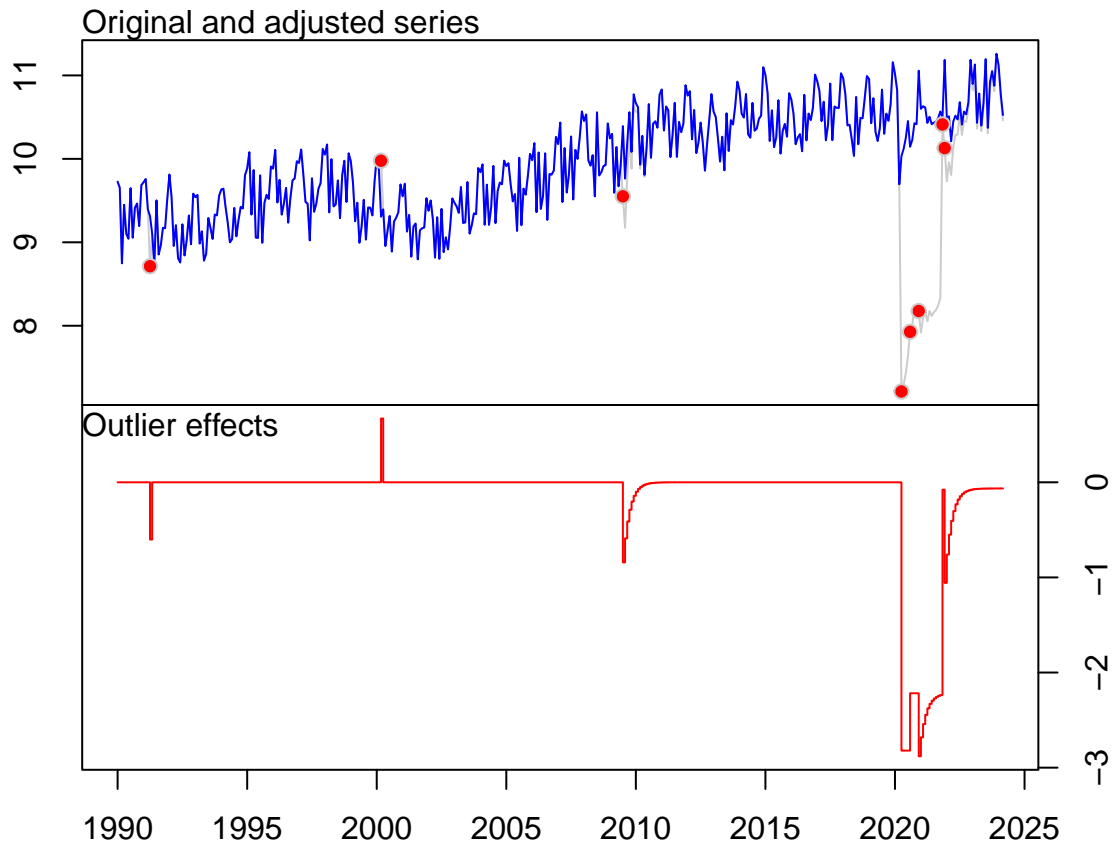


Figure 8: detección de outliers

Raíces unitarias

A continuación nos valeremos de los contrastes de raíz unitaria para analizar si es necesario agregar componentes a nuestro modelo, o, de lo contrario, ya se cuenta con un proceso estacionario, al haber realizado las modificaciones de contener una diferencia estacional y encontrarse en logaritmos

En primer lugar se analiza si es necesario agregar alguna diferencia regular, utilizando el test de Dickey-Fuller aumentado. El resultado determina que no necesaria utilizar ninguna diferencia regular a la serie de tiempo con la que se está trabajando.

```
## [1] 0
```

```
## [1] 0
```

El siguiente paso es analizar si la serie cuenta con un comportamiento tendencial o si es necesario agregar una constante. Para llevar a cabo este análisis se vuelve a realizar el test de raíz unitaria de Dickey-Fuller aumentado. El mensaje que otorga la función utilizada no es muy estetica, por lo que se elige mostrar una tabla con los valores hallados para los distintos contrastes. El valor de “tau3” se omite porque ya se descartó en la parte anterior la presencia de una diferencia regular.

La forma de leer estos números es la siguiente: si el valor del estadístico calculado en el contraste es mayor al respectivo valor crítico al 5%, entonces se rechaza la hipótesis nula correspondiente. Teniendo esto en mente, el estadístico calculado en el primer contraste es mayor al valor crítico, por lo que se rechaza, y se determina

que es necesario agregarle una tendencia al modelo. Y en el segundo contraste, el estadístico también es mayor al valor crítico, entonces se rechaza, y se determina que necesario agregarle una constante al modelo.

	phi2	phi3
Estadístico	6.84	10.26
Valor crítico	4.71	6.30

Estimacion del modelo

Para recapitular, tenemos evidencia que nos sugiere que el modelo a definir debe contener una diferencia estacionaria, ninguna diferencia regular, debe incluir constante, debe incluir un drift, conviene hacer la transformación logarítmica, debe incluir la intervención con los valores atípicos, y además va a tener coeficientes principalmente de la parte MA.

También es importante mencionar que se cuenta con información sobre la fecha en la que cayó carnaval y semana de turismo en el período en el que se está trabajando, y parece razonable pensar que la presencia o ausencia de estos feriados en un mes en específico tiene un efecto considerable en la cantidad de turistas. Así que se integra al modelo esta información en la modalidad de variables indicadoras, esperando ver que sean significativas.

Se trabajará con 2 modelos, uno que incluya semana de turismo y carnaval, y otro que no los incluya. De esta manera se podrá controlar el efecto de estas variables de intervención.

A partir de esto, se pueden intentar diferentes alternativas y combinaciones de parámetros. Luego de haber llevado a cabo la estimación para las variadas posibilidades, se determinó que el modelo que llevaba a las mejores propiedades es el SARIMA(1,0,2)(0,1,1), aunque también requirió algo de trabajo extra para pulir los resultados obtenidos.

El primer comentario que se debe hacer es que todos los coeficientes de los dos modelos son significativos con un nivel de confianza del 95%, lo que es algo muy bueno.

```
##
## z test of coefficients:
##
##      Estimate Std. Error  z value  Pr(>|z|)
## ar1      0.9684051  0.0156284  61.9645 < 2.2e-16 ***
## ma1     -0.8003379  0.0533891 -14.9907 < 2.2e-16 ***
## ma2      0.1277667  0.0549186   2.3265  0.019993 *
## sma1    -0.8427622  0.0322333 -26.1457 < 2.2e-16 ***
## drift    0.0040991  0.0012519   3.2744  0.001059 **
## A016    -0.3914711  0.1516988  -2.5806  0.009863 **
## A0123    0.6300187  0.1501909   4.1948  2.731e-05 ***
## TC235   -0.7578980  0.1269177  -5.9716  2.350e-09 ***
## LS364   -3.0268604  0.1093941 -27.6693 < 2.2e-16 ***
## LS368    0.6898660  0.1103453   6.2519  4.055e-10 ***
## TC372   -0.6593482  0.1320301  -4.9939  5.916e-07 ***
## LS383    2.0955183  0.1255586  16.6896 < 2.2e-16 ***
## TC384   -1.0907226  0.1556557  -7.0073  2.430e-12 ***
## tur      0.0375294  0.0059292   6.3295  2.459e-10 ***
## car      0.1097712  0.0228811   4.7975  1.607e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

##
## z test of coefficients:
##
```

```
##           Estimate Std. Error  z value  Pr(>|z|)
## ar1      0.9664398  0.0164543  58.7348 < 2.2e-16 ***
## ma1     -0.8908803  0.0507218 -17.5641 < 2.2e-16 ***
## ma2      0.2143654  0.0491867   4.3582 1.311e-05 ***
## sma1    -0.8645540  0.0287511 -30.0703 < 2.2e-16 ***
## drift    0.0041217  0.0011230   3.6702 0.0002423 ***
## A016    -0.5741985  0.1577600  -3.6397 0.0002730 ***
## A0123    0.7563842  0.1532007   4.9372 7.925e-07 ***
## TC235   -0.7456073  0.1291343  -5.7739 7.746e-09 ***
## LS364   -2.9910515  0.1106248 -27.0378 < 2.2e-16 ***
## LS368    0.6780285  0.1111825   6.0983 1.072e-09 ***
## TC372   -0.6703720  0.1330074  -5.0401 4.653e-07 ***
## LS383    2.1143226  0.1315919  16.0673 < 2.2e-16 ***
## TC384   -1.1415850  0.1623397  -7.0321 2.035e-12 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Analisis residuos

A la hora de llevar a cabo el diagnóstico del modelo, surgieron resultados favorables, como la falta de correlación de los residuos para cualquier rezago, pero también se evidenció un gran problema, la falta de normalidad de los residuos, que lleva a no poder garantizar que la incorrelación implique independencia.

```
##
##  Shapiro-Wilk normality test
##
## data:  residuos1
## W = 0.99215, p-value = 0.02915
##
##  Jarque Bera Test
##
## data:  residuos1
## X-squared = 17.882, df = 2, p-value = 0.0001309
##
##  Box-Ljung test
##
## data:  residuos1
## X-squared = 19.357, df = 25, p-value = 0.7797
##
##  Shapiro-Wilk normality test
##
## data:  residuos1.1
## W = 0.98579, p-value = 0.0004715
##
##  Jarque Bera Test
##
## data:  residuos1.1
## X-squared = 43.478, df = 2, p-value = 3.621e-10
##
##  Box-Ljung test
##
## data:  residuos1.1
```

```
## X-squared = 28.299, df = 25, p-value = 0.2943
```

Una posible estrategia para intentar solucionar la falta de normalidad de los residuos es estudiar la presencia de valores atípicos, por lo que se propuso estudiar si hay algún residuo que sea mayor al resto y pueda estar distorsionando los resultados. Se puede ver en el gráfico como el residuo del inicio de la pandemia es excesivamente grande, por lo que se decide intervenir manualmente el modelo en ese punto

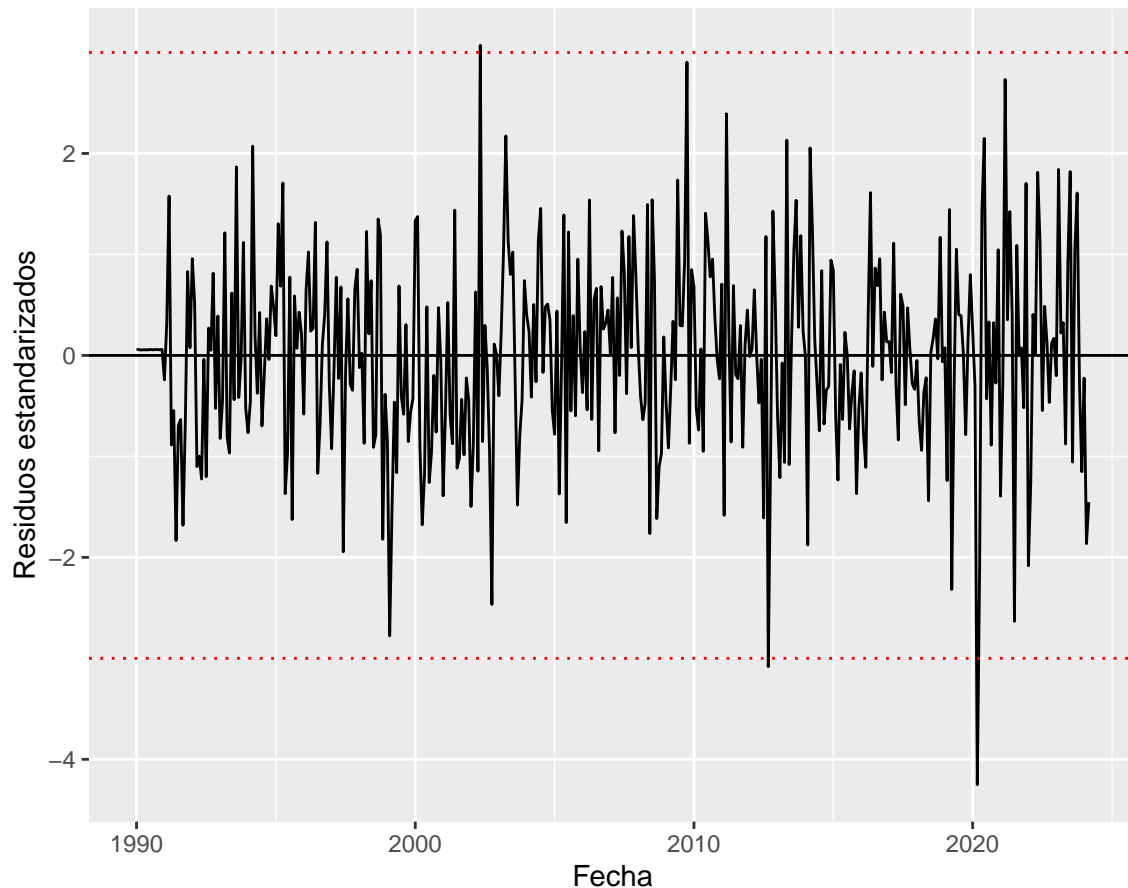


Figure 9: Residuos de un modelo SARIMA(0,1,2)(0,1,1) para el logaritmo del turismo mensual brasileiro, incluyendo semana de turismo y carnaval

```
## [1] 2020.167
```

Reestimación Modelo

Se vuelven a correr los mismos modelos, pero esta vez con la nueva variable de intervención creada, que intentará corregir la falta de normalidad en los residuos.

En ambos casos todos los parametros siguen siendo significativos, incluidos la nueva intervención agregada, por lo que puede ser de esperar al menos mejorra los resultados anteriores

Modelo 2 con carnaval y semana de turismo

```
##
```

```
## z test of coefficients:
##
##      Estimate Std. Error  z value  Pr(>|z|)
## ar1      0.9645142  0.0172009   56.0735 < 2.2e-16 ***
## ma1     -0.8290776  0.0521171  -15.9080 < 2.2e-16 ***
## ma2      0.1928932  0.0560566    3.4410 0.0005795 ***
## sma1    -0.8277152  0.0342408  -24.1734 < 2.2e-16 ***
## drift    0.0045354  0.0013013    3.4853 0.0004916 ***
## A016    -0.3939642  0.1449405   -2.7181 0.0065656 **
## A0123    0.6406206  0.1429988    4.4799 7.468e-06 ***
## TC235   -0.7638223  0.1207273   -6.3268 2.502e-10 ***
## LS364   -3.2262841  0.1125749  -28.6590 < 2.2e-16 ***
## LS368    0.6684866  0.1061040    6.3003 2.971e-10 ***
## TC372   -0.6654919  0.1239329   -5.3698 7.883e-08 ***
## LS383    2.1144204  0.1272365   16.6180 < 2.2e-16 ***
## TC384   -1.1056761  0.1509729   -7.3237 2.413e-13 ***
## A0      -0.7856006  0.1521170   -5.1644 2.411e-07 ***
## tur      0.0338824  0.0057764    5.8657 4.473e-09 ***
## car      0.1008282  0.0223102    4.5194 6.202e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Modelo 2.1 sin carnaval y semana de turismo

```
##
## z test of coefficients:
##
##      Estimate Std. Error  z value  Pr(>|z|)
## ar1      0.9624580  0.0183030   52.5848 < 2.2e-16 ***
## ma1     -0.9059758  0.0498342  -18.1798 < 2.2e-16 ***
## ma2      0.2663551  0.0503683    5.2881 1.236e-07 ***
## sma1    -0.8527946  0.0302734  -28.1698 < 2.2e-16 ***
## drift    0.0045615  0.0011553    3.9483 7.871e-05 ***
## A016    -0.5567811  0.1493413   -3.7282 0.0001928 ***
## A0123    0.7532435  0.1445648    5.2104 1.884e-07 ***
## TC235   -0.7549948  0.1229446   -6.1409 8.204e-10 ***
## LS364   -3.2049268  0.1127160  -28.4336 < 2.2e-16 ***
## LS368    0.6546179  0.1063082    6.1577 7.379e-10 ***
## TC372   -0.6708539  0.1246633   -5.3813 7.394e-08 ***
## LS383    2.1398352  0.1334199   16.0384 < 2.2e-16 ***
## TC384   -1.1550298  0.1570013   -7.3568 1.883e-13 ***
## A0      -0.9051087  0.1569803   -5.7657 8.130e-09 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Diagnostico del modelo

Análisis gráfico residuos

Ya no hay presencia de residuos excesivamente grande. Se podría intervenir alguno más que cruza por poco el umbral de los tres desvíos, pero veremos los resultados de los demás tests antes de tomar la decisión de realizar otra intervención. Mirando las FAC y FACP, sucede algo raro en los residuos 9 y 17, pero a falta de una razón aparente que podría explicar esto, no se toma mayor acción al respecto

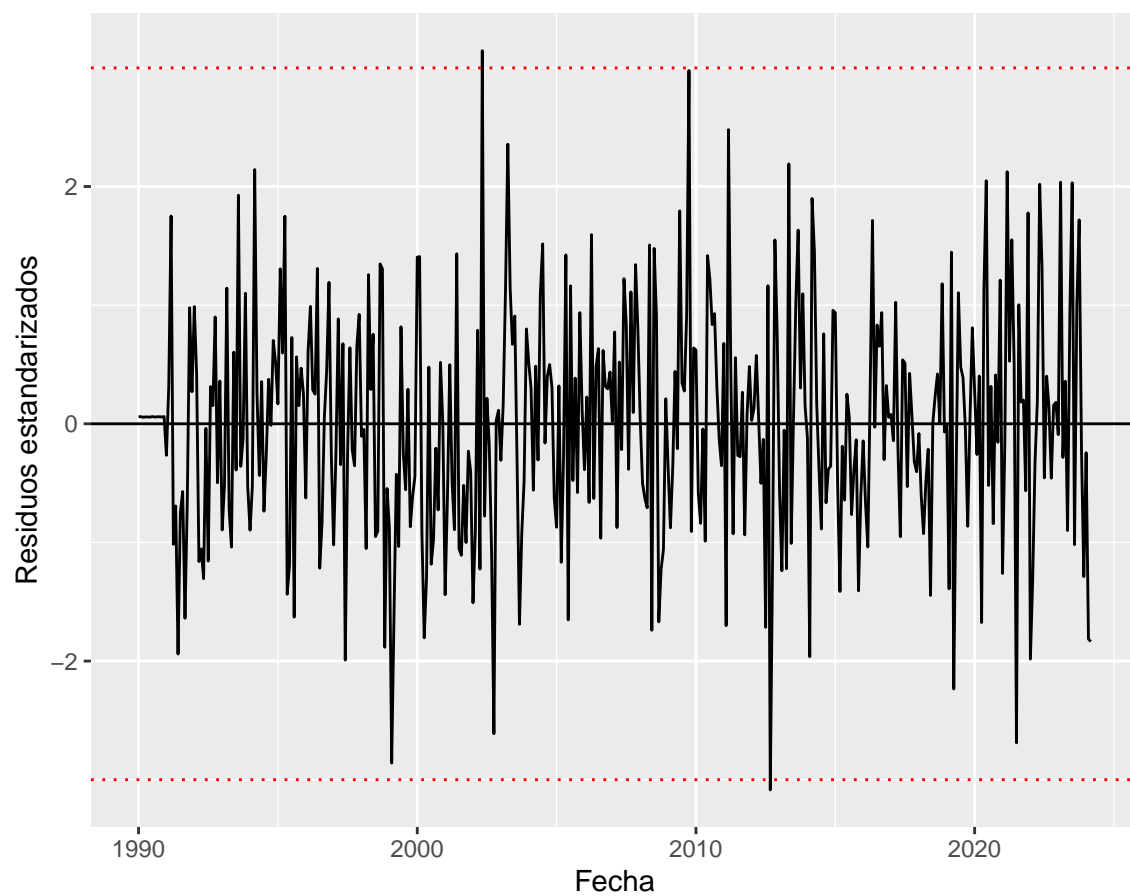


Figure 10: Residuos de un modelo SARIMA(1,0,2)(0,1,1) para el logaritmo del turismo mensual brasileiro, incluyendo semana de turismo y carnaval

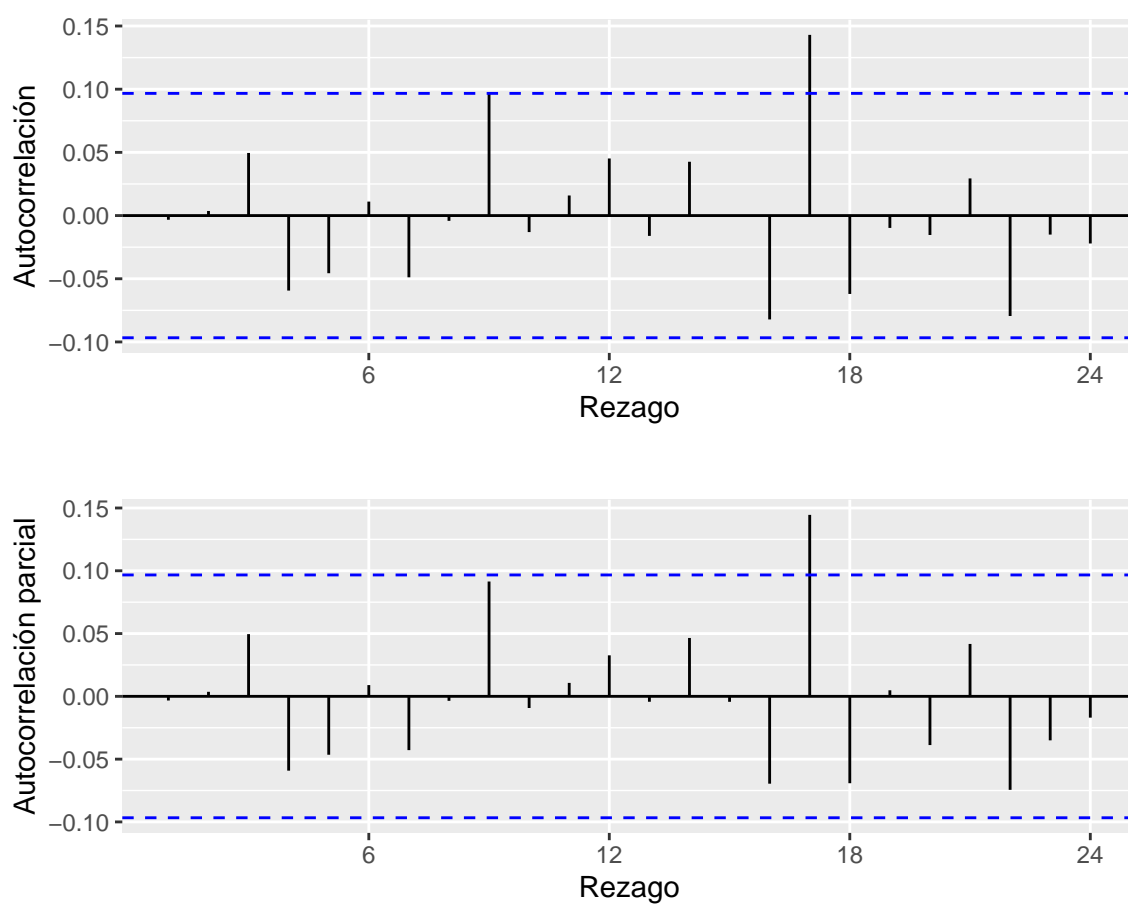


Figure 11: Funciones de Autocorrelacion y Autocorrelación parcial estimadas de los residuos de un modelo SARIMA(1,0,2)(0,1,1) para el logaritmo del turismo mensual proveniente de Brasil entre 1990 y 2024,incluyendo semana de turismo y carnaval

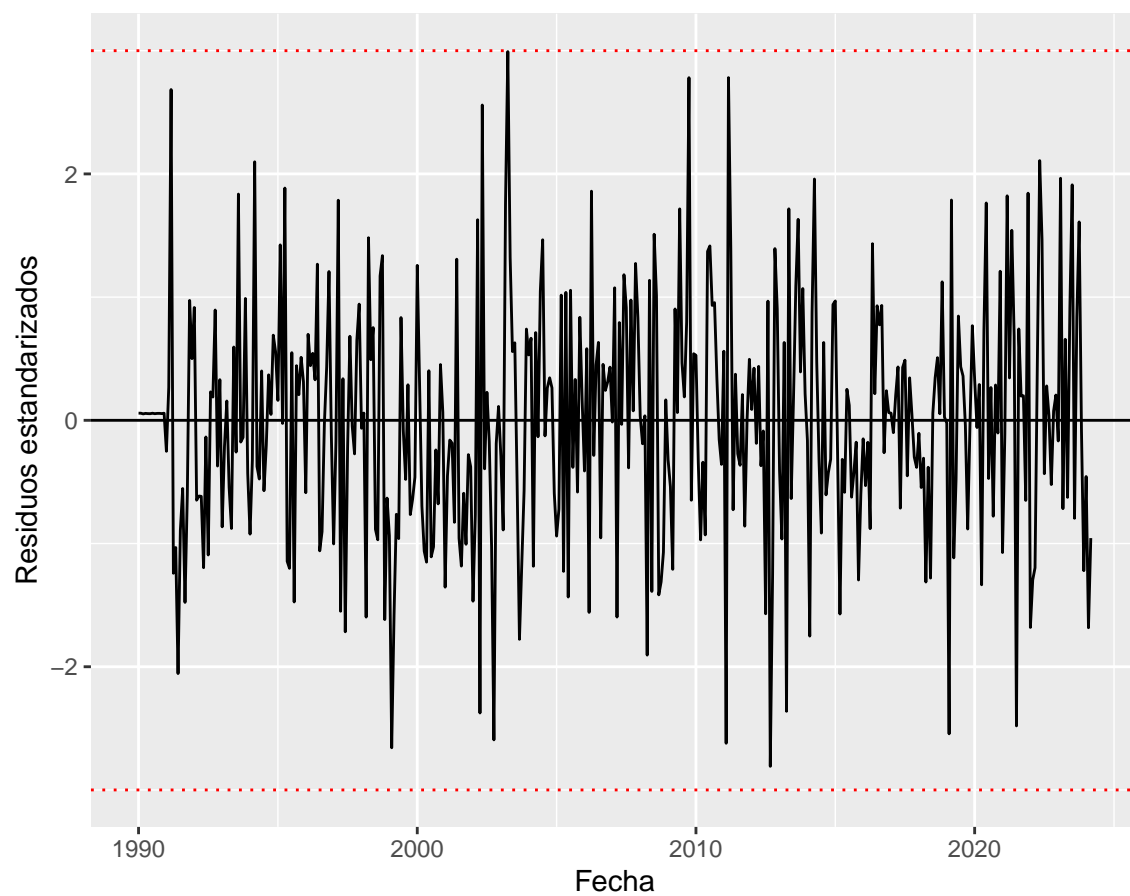


Figure 12: Residuos de un modelo SARIMA(1,0,2)(0,1,1) para el logaritmo del turismo mensual brasileiro

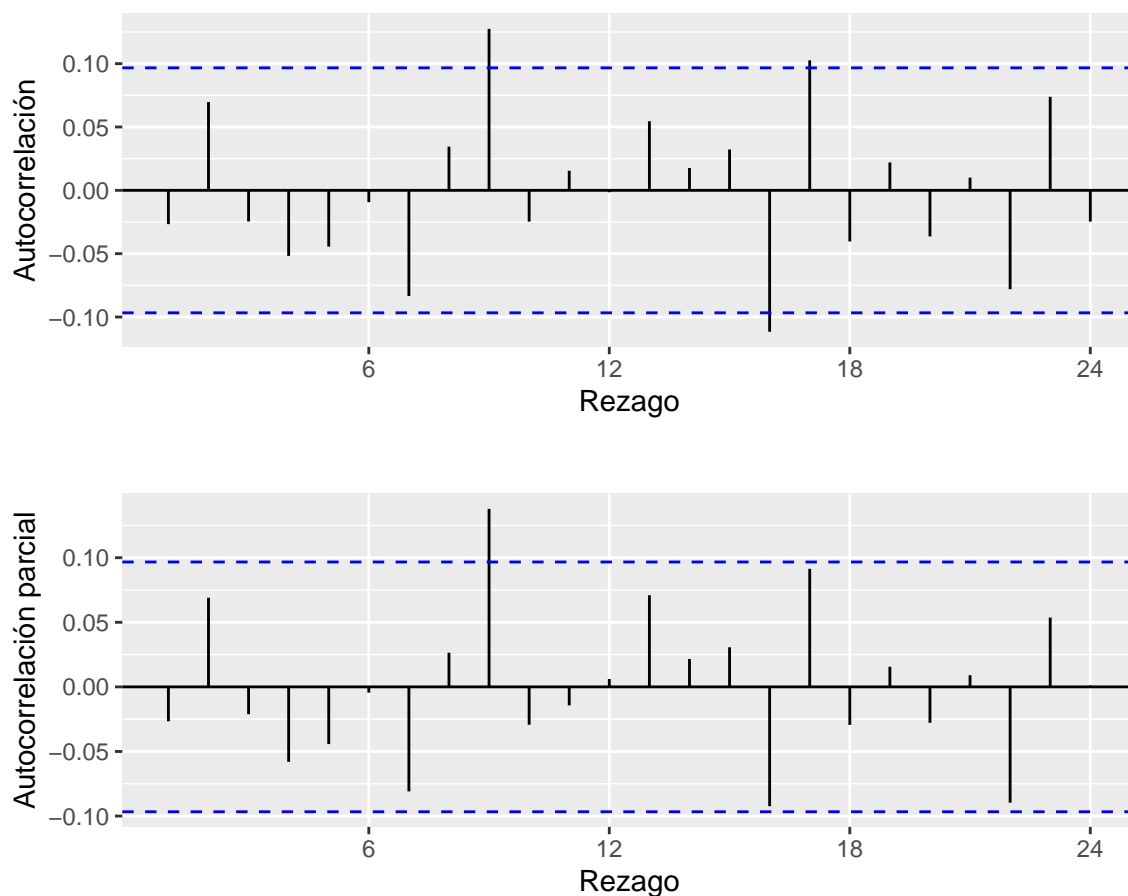


Figure 13: Funciones de Autocorrelacion y Autocorrelación parcial estimadas de los residuos de un modelo SARIMA(1,0,2)(0,1,1) para el logaritmo del turismo mensual proveniente de Brasil entre 1990 y 2024

contraste autocorrelacion de los residuos

Más allá del análisis gráfico, se realiza el contraste de Ljung-Box para evaluar la incorrelación de los residuos. El objetivo del contraste es no rechazar. Entonces, se puede confirmar que ninguno de los modelos presenta correlación entre los residuos, pero, además, se puede decir que el modelo que incluye a semana de turismo y carnaval rechaza con más margen la hipótesis nula, por lo que hay todavía más evidencias para confirmar la falta de correlación

```
##
## Box-Ljung test
##
## data:  residuos2
## X-squared = 28.519, df = 25, p-value = 0.2845

##
## Box-Ljung test
##
## data:  residuos2.1
## X-squared = 33.93, df = 25, p-value = 0.1094
```

normalidad para el modelo 2

Para estudiar la normalidad, en primer lugar se presentan dos opciones gráficas que permiten detectar de algún comportamiento inesperado o simplemente ver que se comporta correctamente.

Ambos modelos presentan un comportamiento similar, y hay razones para pensar que esta vez sí los residuos se distribuyen de manera normal.

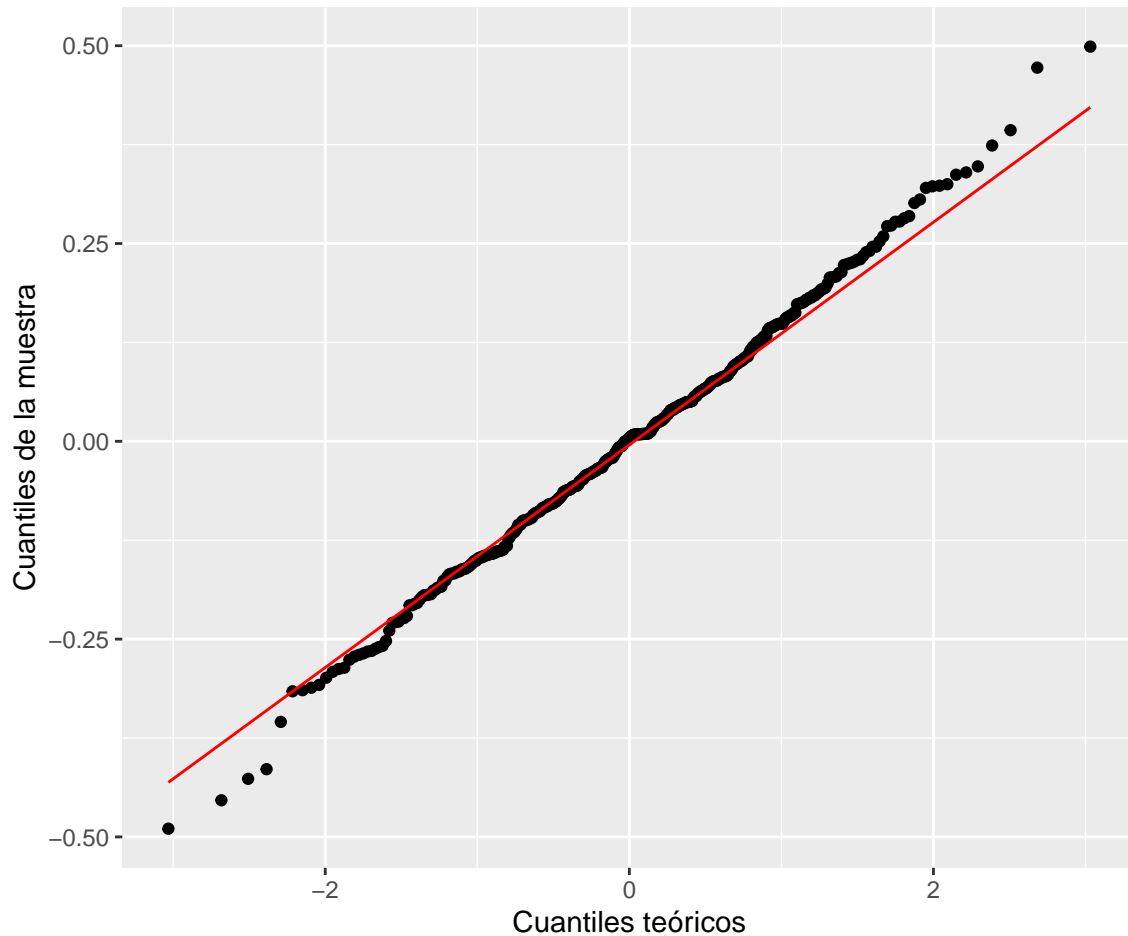


Figure 14: QQ-plot de los residuos de un modelo SARIMA(1,0,2)(0,1,1) para el logaritmo del turismo mensual proveniente de Brasil entre 1990 y 2024,incluyendo semana de turismo y carnaval

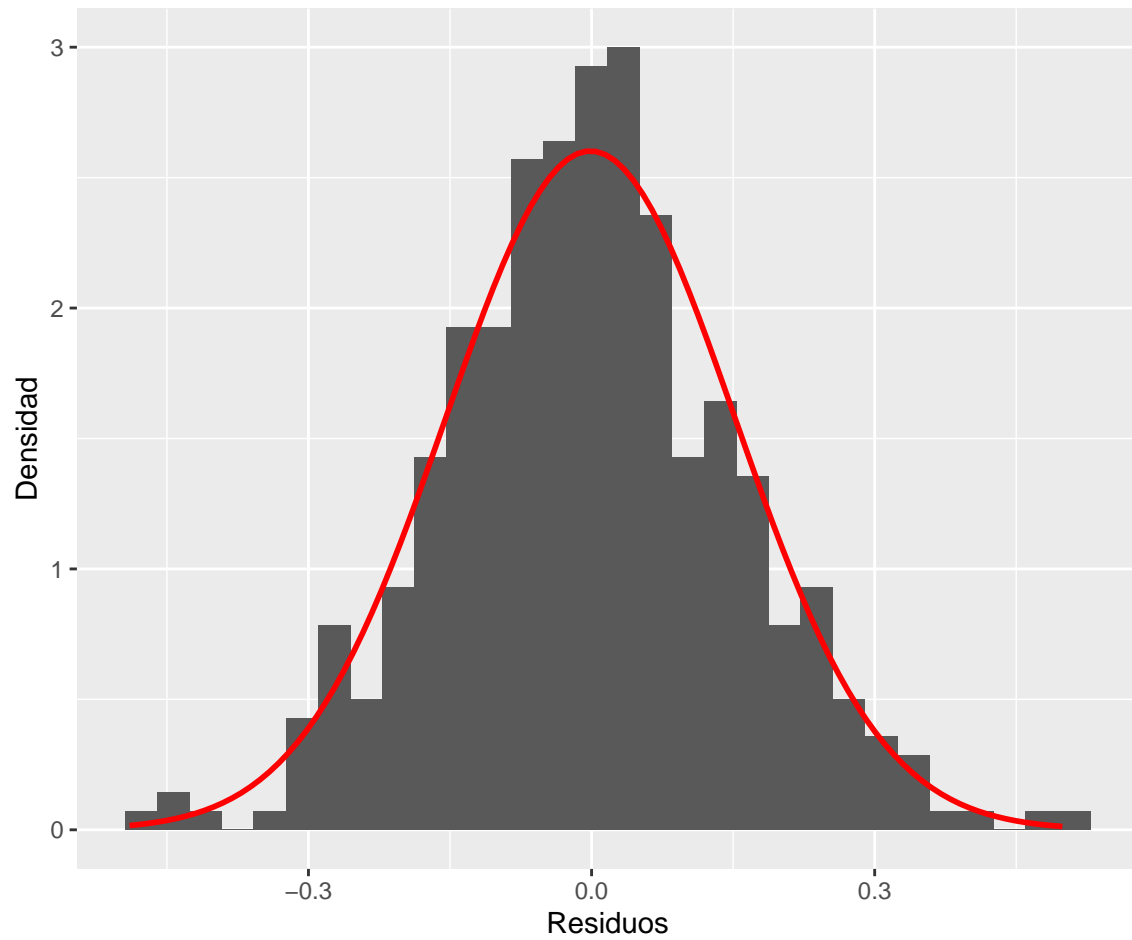


Figure 15: Histograma de los residuos de un modelo SARIMA(1,0,2)(0,1,1) para el logaritmo del turismo mensual proveniente de Brasil entre 1990 y 2024,incluyendo semana de turismo y carnaval

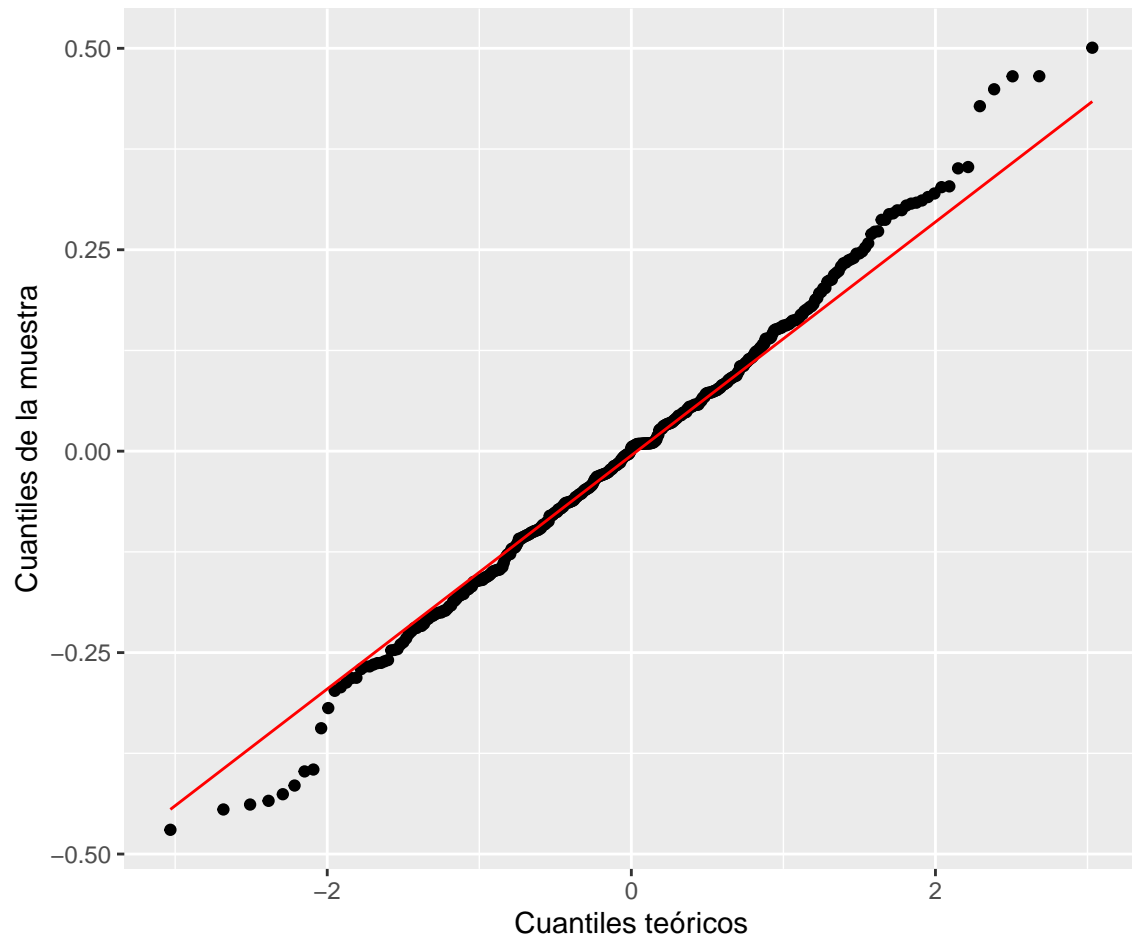


Figure 16: QQ-plot de los residuos de un modelo SARIMA(1,0,2)(0,1,1) para el logaritmo del turismo mensual proveniente de Brasil entre 1990 y 2024

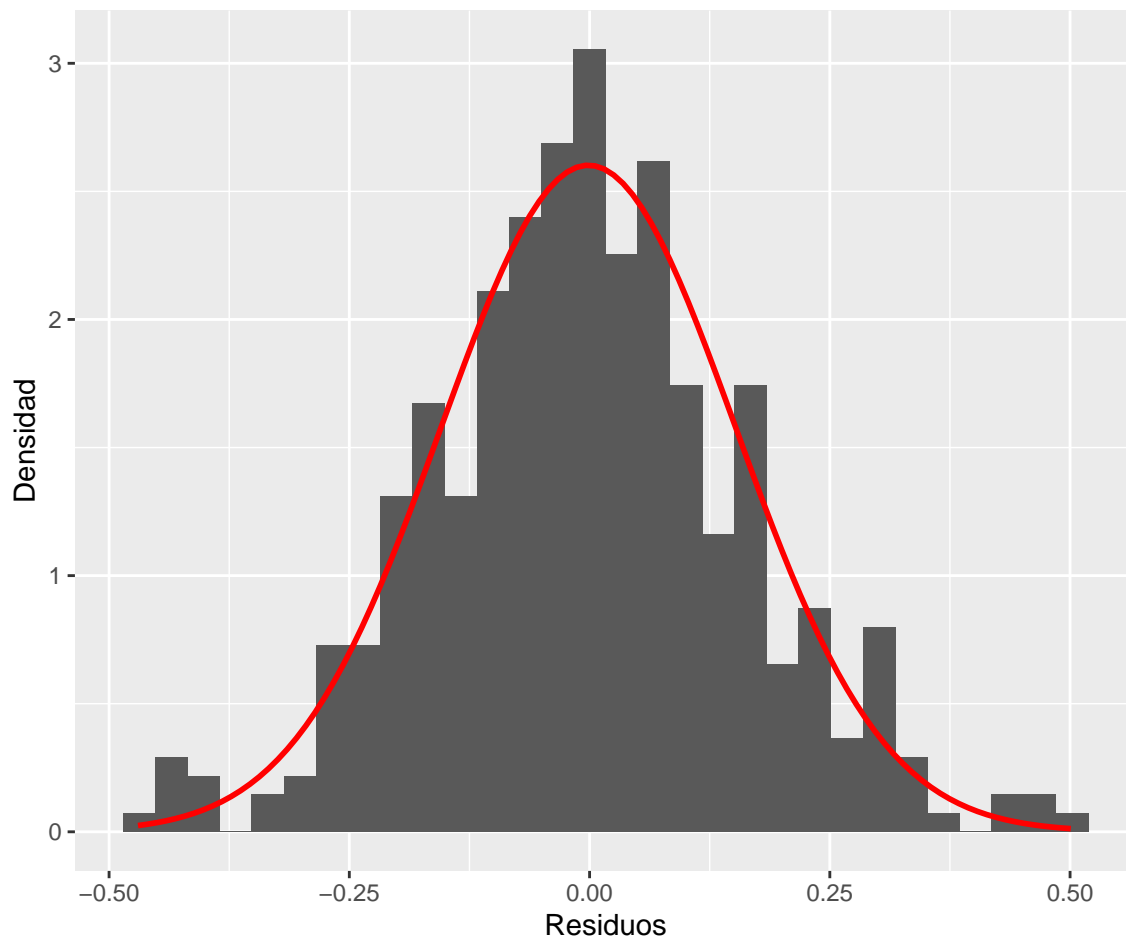


Figure 17: Histogramas de los residuos de un modelo SARIMA(1,0,2)(0,1,1) para el logaritmo del turismo mensual proveniente de Brasil entre 1990 y 2024

contraste normalidad

Para confirmar si se está en presencia de normalidad en los residuos, se llevan a cabo dos tests que evalúan la normalidad. Estos son el test de Shapiro-Wilk y el test de Jarque-Bera, y en ambos el objetivo es no rechazar.

Viendo los resultados, es impresionante como una simple intervención en una observaciones pudo haber cambiado tanto, siendo que ahora todos los tests llevan a rechazar, y con un interesante margen.

```
##
##  Shapiro-Wilk normality test
##
## data:  residuos2
## W = 0.99727, p-value = 0.7319
##
##  Jarque Bera Test
##
## data:  residuos2
## X-squared = 1.7947, df = 2, p-value = 0.4076
##
##  Shapiro-Wilk normality test
```

```
##
## data:  residuos2.1
## W = 0.99496, p-value = 0.2012

##
## Jarque Bera Test
##
## data:  residuos2.1
## X-squared = 2.3637, df = 2, p-value = 0.3067
```

Predicción

Una importante propiedad que tiene que presentar un modelo es la capacidad de predecir con precisión. Al no contar con datos para más adelante en el tiempo, se realiza la predicción sin realmente saber si será buena o no. De cualquier manera, en ambos modelos se realizará una predicción 12 meses hacia adelante.

Modelo 2

```
##          Point Forecast      Lo 51      Hi 51      Lo 54      Hi 54      Lo 57      Hi 57
## Apr 2024      33276.93 29454.86 36666.52 29228.94 36949.92 28996.44 37246.19
## May 2024      32199.89 28466.55 35507.17 28246.23 35784.12 28019.51 36073.67
## Jun 2024      33731.36 29617.20 37354.00 29376.53 37660.03 29128.96 37980.10
## Jul 2024      49828.75 43483.70 55385.79 43115.40 55858.89 42736.70 56353.88
## Aug 2024      30882.26 26801.15 34439.36 26565.92 34744.31 26324.11 35063.47
## Sep 2024      40029.93 34566.36 44771.42 34253.42 45180.45 33931.82 45608.67
## Oct 2024      42141.91 36224.79 47256.99 35887.79 47700.76 35541.55 48165.45
## Nov 2024      43392.52 37145.33 48773.84 36791.35 49243.10 36427.76 49734.59
## Dec 2024      71171.76 60694.63 80167.73 60103.75 80955.85 59496.96 81781.50
## Jan 2025      56152.42 47720.32 63371.25 47246.80 64006.37 46760.62 64671.86
## Feb 2025      41654.51 35286.95 47091.28 34930.76 47571.46 34565.12 48074.70
## Mar 2025      32949.04 27830.74 37308.45 27545.45 37694.85 27252.64 38099.87
##          Lo 60      Hi 60      Lo 63      Hi 63      Lo 66      Hi 66      Lo 69
## Apr 2024 28756.29 37557.25 28507.16 37885.47 28247.46 38233.77 27975.20
## May 2024 27785.33 36377.71 27542.43 36698.53 27289.24 37039.01 27023.82
## Jun 2024 28873.36 38316.32 28608.35 38671.26 28332.24 39048.12 28042.94
## Jul 2024 42345.85 56874.02 41940.76 57423.34 41518.88 58006.83 41077.02
## Aug 2024 26074.63 35398.95 25816.15 35753.38 25547.05 36129.98 25265.31
## Sep 2024 33600.11 46058.93 33256.55 46534.75 32898.99 47040.51 32524.76
## Oct 2024 35184.53 48654.19 34814.84 49170.83 34430.21 49720.13 34027.78
## Nov 2024 36052.94 50251.66 35664.92 50798.37 35261.32 51379.81 34839.16
## Dec 2024 58871.55 82650.29 58224.28 83569.10 57551.17 84546.50 56847.28
## Jan 2025 46259.63 65372.26 45741.23 66113.15 45202.25 66901.45 44638.76
## Feb 2025 34188.39 48604.44 33798.66 49164.90 33393.54 49761.35 32970.08
## Mar 2025 26951.00 38526.28 26639.00 38977.50 26314.75 39457.79 25975.88
##          Hi 69      Lo 72      Hi 72      Lo 75      Hi 75      Lo 78      Hi 78
## Apr 2024 38605.88 27687.81 39006.60 27381.93 39442.33 27053.04 39921.84
## May 2024 37402.80 26743.68 37794.60 26445.56 38220.66 26125.02 38689.60
## Jun 2024 39450.96 27737.74 39885.05 27413.12 40357.36 27064.31 40877.49
## Jul 2024 58630.80 40611.10 59303.47 40115.76 60035.72 39583.80 60842.53
## Aug 2024 36532.88 24968.35 36967.38 24652.79 37440.58 24314.04 37962.21
## Sep 2024 47581.76 32130.45 48165.69 31711.60 48801.87 31262.17 49503.45
## Oct 2024 50308.15 33603.89 50942.75 33153.77 51634.38 32670.98 52397.40
## Nov 2024 52002.41 34394.62 52674.52 33922.73 53407.26 33416.75 54215.92
## Dec 2024 85593.37 56106.29 86723.80 55319.93 87956.55 54477.04 89317.45
```

##	Jan	2025	67745.97	44045.72	68658.12	43416.53	69653.11	42742.30	70751.84
##	Feb	2025	50400.47	32524.51	51090.94	32051.90	51844.28	31545.58	52676.40
##	Mar	2025	39972.54	25619.39	40528.75	25241.35	41135.75	24836.45	41806.38
##			Lo 81	Hi 81	Lo 84	Hi 84	Lo 87	Hi 87	Lo 90
##	Apr	2024	26694.73	40457.68	26297.59	41068.68	25846.67	41785.15	25316.11
##	May	2024	25775.88	39213.67	25388.92	39811.32	24949.65	40512.26	24432.87
##	Jun	2024	26684.60	41459.15	26264.08	42122.96	25787.09	42902.12	25226.49
##	Jul	2024	39005.05	61745.31	38364.49	62776.25	37638.45	63987.20	36785.85
##	Aug	2024	23945.69	38546.18	23538.23	39213.42	23076.70	39997.69	22535.13
##	Sep	2024	30773.68	50289.24	30233.62	51187.55	29622.24	52244.02	28905.34
##	Oct	2024	32146.45	53252.37	31566.81	54230.21	30910.96	55380.82	30142.40
##	Nov	2024	32867.24	55122.36	32260.25	56159.51	31573.79	57380.51	30769.80
##	Dec	2024	53561.94	90843.44	52551.50	92590.14	51409.26	94647.35	50072.16
##	Jan	2025	42010.53	71984.24	41202.82	73395.38	40290.10	75058.04	39222.18
##	Feb	2025	30996.22	53610.00	30390.03	54679.36	29705.30	55939.77	28904.46
##	Mar	2025	24397.23	42559.00	23912.73	43421.31	23365.62	44438.01	22726.00
##			Hi 90	Lo 93	Hi 93	Lo 96	Hi 96	Lo 99	Hi 99
##	Apr	2024	42660.86	24654.04	43806.49	23726.17	45519.66	21840.41	49449.95
##	May	2024	41369.13	23788.15	42490.34	22884.85	44167.51	21050.03	48017.35
##	Jun	2024	43855.53	24527.91	45104.56	23550.77	46976.01	21571.99	51285.07
##	Jul	2024	65470.24	35724.57	67415.20	34242.22	70333.61	31248.48	77071.86
##	Aug	2024	40958.92	21861.64	42220.73	20922.18	44116.57	19029.43	48504.59
##	Sep	2024	53539.77	28014.56	55242.16	26773.45	57802.97	24278.41	63743.25
##	Oct	2024	56792.91	29188.18	58649.59	27860.07	61445.45	25195.38	67943.99
##	Nov	2024	58879.81	29772.29	60852.55	28385.28	63826.05	25607.34	70750.02
##	Dec	2024	97174.77	48414.26	100502.42	46111.00	105522.55	41505.47	117231.53
##	Jan	2025	77101.68	37898.82	79793.94	36061.77	83858.78	32393.90	93353.89
##	Feb	2025	57489.66	27912.59	59532.55	26536.71	62619.19	23793.32	69839.25
##	Mar	2025	45688.72	21934.19	47338.05	20836.56	49831.73	18650.66	55672.13

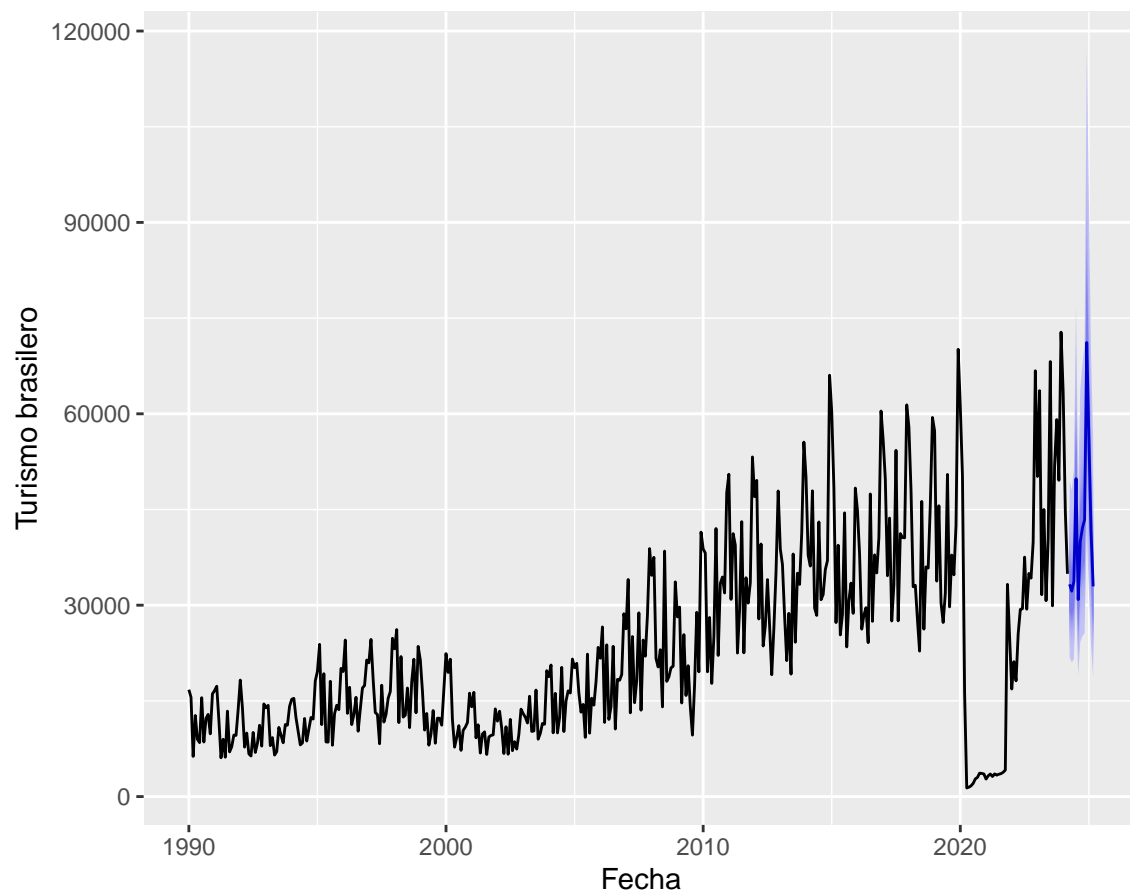


Figure 18: Predicciones a 12 pasos de un modelo SARIMA(1,0,2)(0,1,1) para el logaritmo del turismo mensual proveniente de Brasil entre 1990 y 2024,incluyendo semana de turismo y carnaval

Modelo 2.1

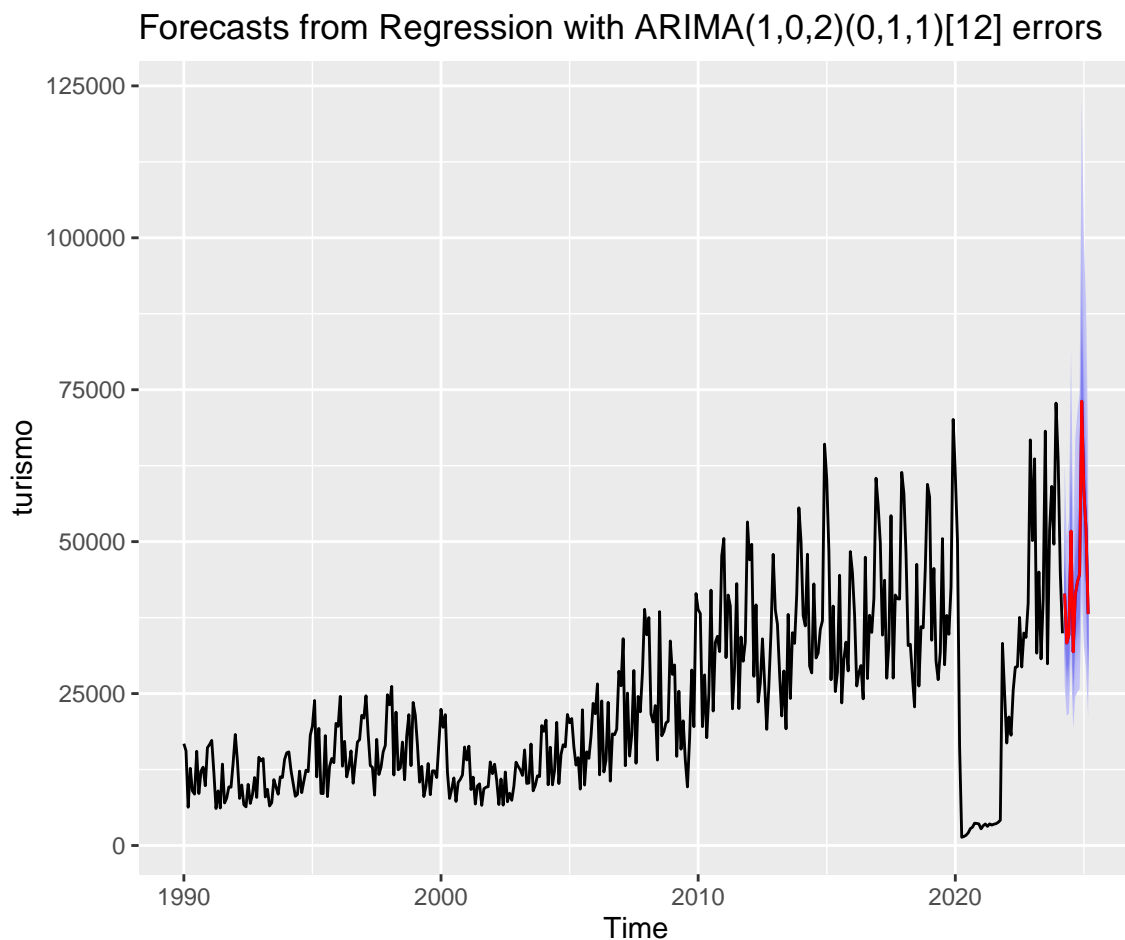


Figure 19: Predicciones a 12 pasos de un modelo SARIMA(1,0,2)(0,1,1) para el logaritmo del turismo mensual proveniente de Brasil entre 1990 y 2024

```
## $x
## [1] "Fecha"
##
## $y
## [1] "Turismo brasilero"
##
## $title
## [1] ""
##
## attr(,"class")
## [1] "labels"
```

Validacion de las predicciones

Una mejor forma de evaluar las predicciones es realizarla en puntos donde si se conoce el valor real que va a tomar. La forma de llevar esto a cabo es mediante una muestra de entrenamiento y una muestra de testeo. Esto significa que se utilizará una parte de la muestra para estimar el modelo, y se realiza a la predicción en

la parte que no formó parte de la estimación. De esta manera se asegura no sufrir sobreajuste, que es cuando la predicción está sesgada por predecir en los mismos valores que se utilizaron para la estimación.

En este caso se utilizará los últimos dos años que contienen los datos para testear las predicciones.

Modelo 2

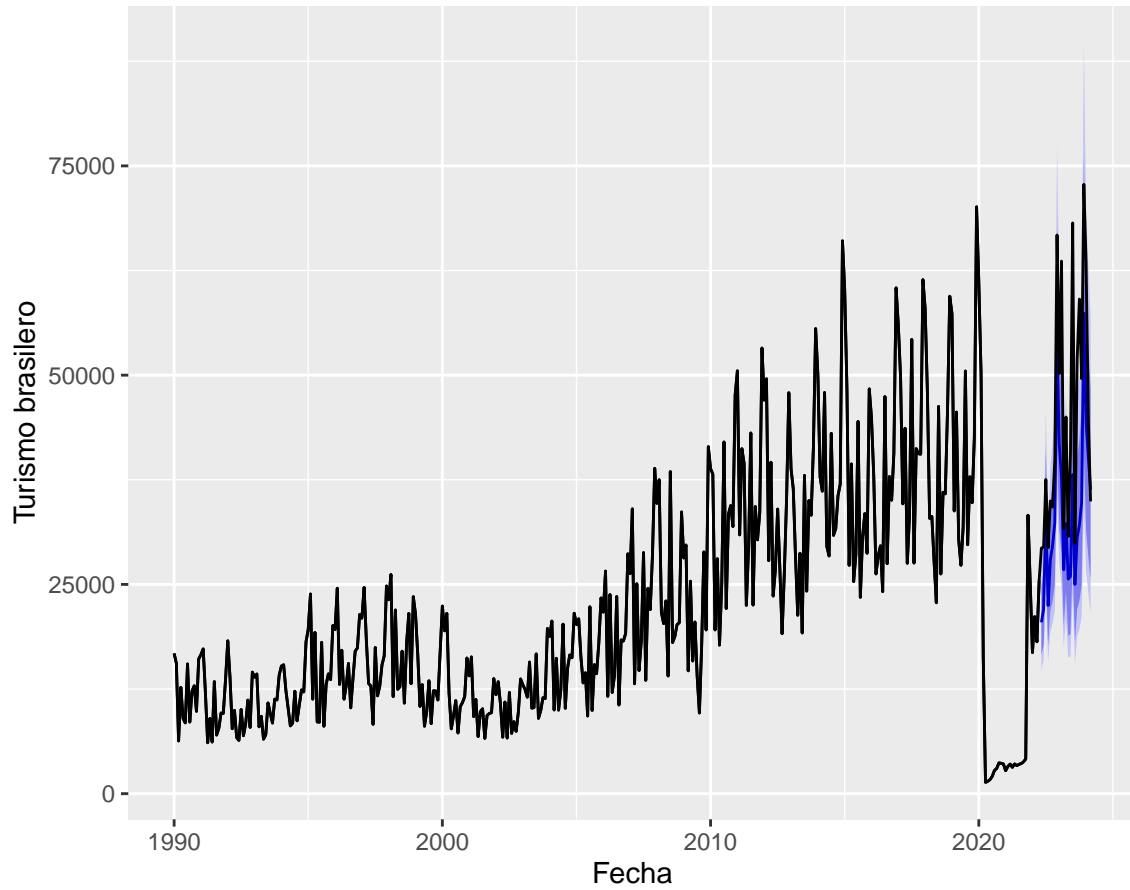


Figure 20: Predicciones en el conjunto de prueba de un modelo SARIMA(1,0,2)(0,1,1) para el logaritmo del turismo mensual proveniente de Brasil entre 1990 y 2024,incluyendo semana de turismo y carnaval

Modelo 2.1

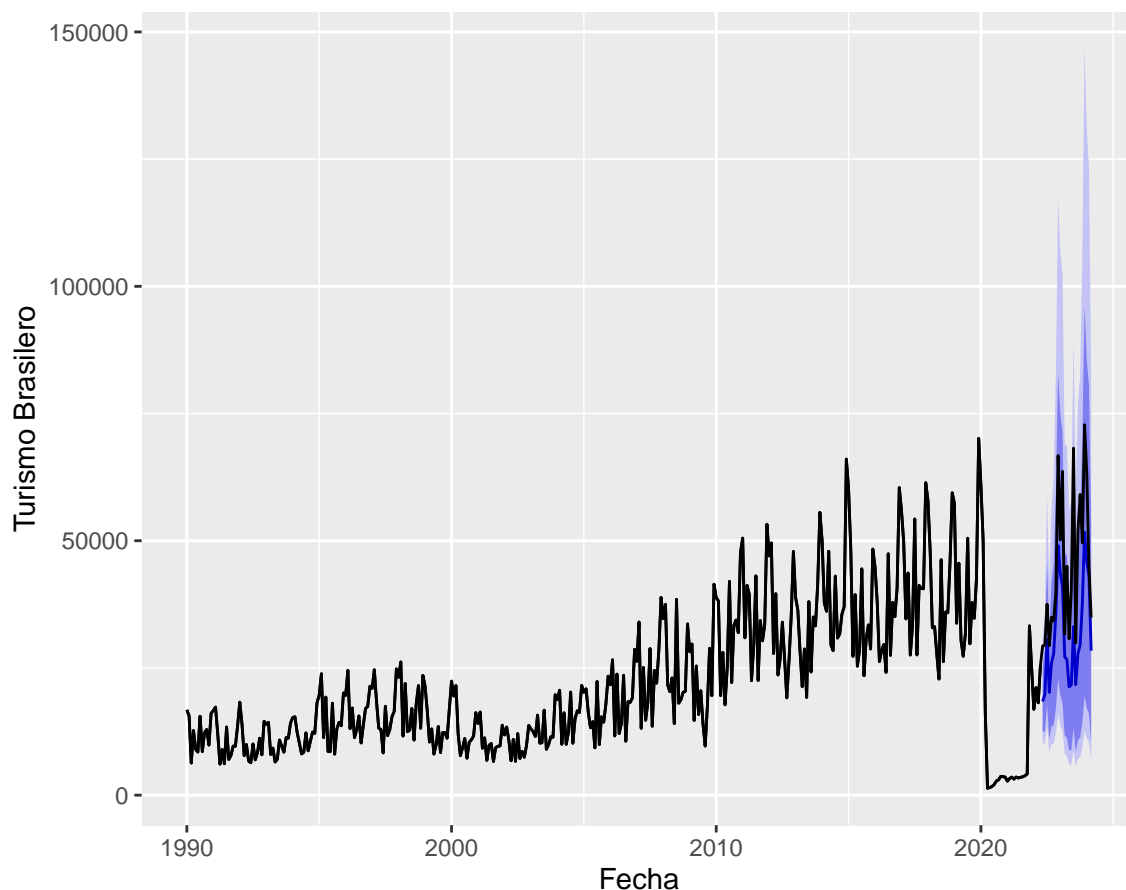


Figure 21: Predicciones en el conjunto de prueba de un modelo SARIMA(1,0,2)(0,1,1) para el logaritmo del turismo mensual proveniente de Brasil entre 1990 y 2024

Comparación

Al comparar el desempeño predictivo de cada uno de los modelos, se nota claramente que el modelo que incluye a semana de turismo y carnaval obtuvo mucho mejores resultados, y esto en realidad era lo que esperaba. Si se incluya más información relevante al modelo, uno esperaría que termine obteniendo un mejor modelo.

##	ME	RMSE	MAE	MPE	MAPE	MASE
## Training set	-188.624	3388.367	2378.468	-2.508378	12.12852	0.4118202
## Test set	11590.243	14153.175	11712.065	23.898447	24.24734	2.0278867
##	ACF1	Theil's U				
## Training set	0.005610179	NA				
## Test set	0.079968863	0.9087159				

##	ME	RMSE	MAE	MPE	MAPE	MASE
## Training set	-535.1603	5694.632	3229.486	-8.964363	20.01551	0.5591697
## Test set	13475.9436	15940.370	13475.944	28.719159	28.71916	2.3332936
##	ACF1	Theil's U				
## Training set	-0.09565779	NA				

```
## Test set          0.01327061  1.066473
```

Interpretación y conclusiones

Luego de ir probando distintos modelos, podemos llegar a la conclusión de que el modelo que hace que la serie sea tendencialmente estacionaria (TS) y supera el supuesto de normalidad es el modelo que contiene un drift, una constante, y además a los outliers, la intervención de la pandemia (AO) y las variables de carnaval y turismo. Al realizar la predicción a 12 pasos, podemos ver que a un 95% de confianza, la estimación a un punto es buena.

En base a los resultados hallados respecto al cumplimiento de todos los diagnósticos, la significación de todos los parámetros, y el haber conseguido una buena precisión a la hora de predecir, sentimos que estamos en condiciones de asegurar que el modelo llevado a cabo, no solo es válido, sino que también es útil. De esta manera podemos concluir que la cantidad de turistas provenientes de Brasil depende de: la cantidad de turistas del mes anterior, de los errores producidos en los dos meses anteriores, de la cantidad de turistas del mismo mes del año anterior, y del error producido en ese mismo momento. Pero más allá de eso, se logró detectar que, independientemente de todos esos valores, hay un piso de turistas brasileños que vendrán al país cada mes, indicado por la constante, y, además, hay una tendencia al alza en la cantidad total, por lo que el crecimiento será de mayor magnitud que la esperada solo por el efecto de los períodos que inciden en el momento de análisis.

Anexo Se agrega el tema de descomposición visto al final de curso con el fin de estudiar los componentes no observables de tendencias, ciclos y estacionalidad para validar los resultados alcanzados.

Descomposición

Adicional: Tendencia-ciclo, estacionalidad y componente irregular

El objetivo es estudiar los componentes no observables mediante la metodología X13. La función `x13_spec` de la librería `RJDemetra` tiene incorporado el efecto calendario como también detecta automáticamente los outliers como los AO. En el tercer gráfico podemos confirmar una clara estacionalidad en la serie como ya fue estudiado anteriormente. Por último estamos interesados en la predicción de los componentes no observables 3 años hacia adelante (hasta marzo del 2027).

```
spec_x13 <- x13_spec(spec = "RSA5c", x11.fcasts = -3, easter.duration = 7, transform.function = "Log",
                    outlier.enabled = TRUE, outlier.ao = TRUE) ## tiene en cuenta turismo pero no ca

# Descomposición X13
modelo_x13_turismo <- x13(turismo, spec = spec_x13)

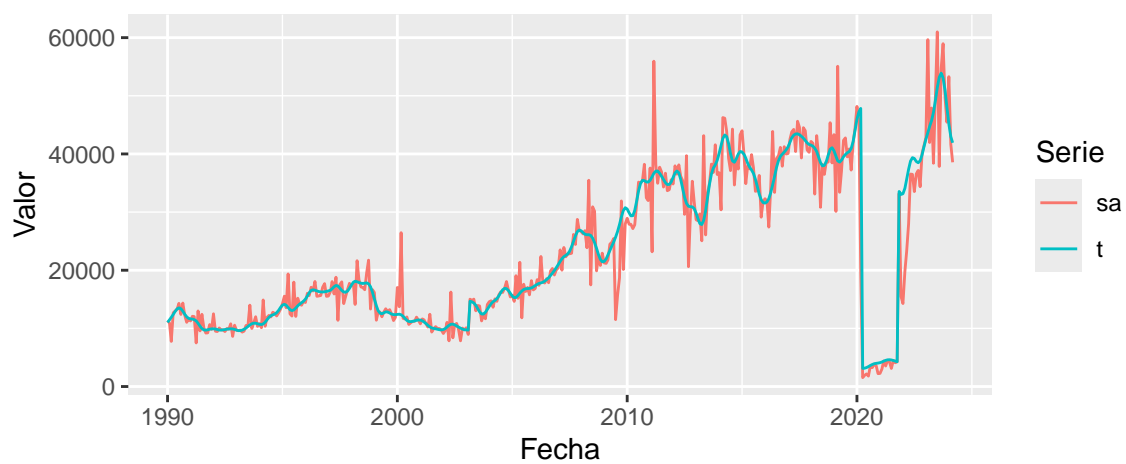
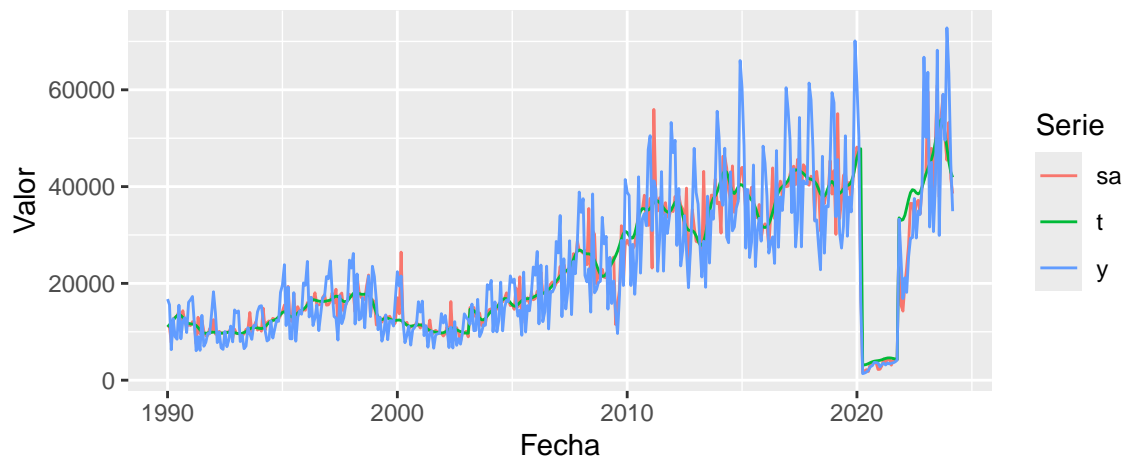
summary(modelo_x13_turismo$regarima)

## y = regression model + arima (1, 1, 1, 0, 1, 1)
##
## Model: RegARIMA - X13
## Estimation span: from 1-1990 to 3-2024
## Log-transformation: yes
## Regression model: no mean, no trading days effect, no leap year effect, Easter effect, outliers(8)
##
## Coefficients:
```

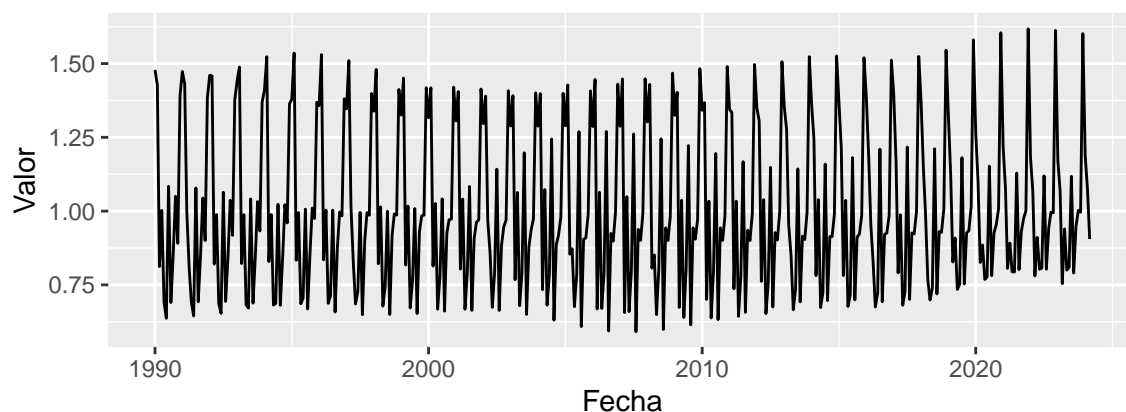
```

## ARIMA:
##           Estimate Std. Error  T-stat Pr(>|t|)
## Phi(1)      0.32140    0.06231   5.158 3.94e-07 ***
## Theta(1)    -0.60789    0.05287 -11.497 < 2e-16 ***
## BTheta(1)   -0.84320    0.03074 -27.430 < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Regression model:
##           Estimate Std. Error  T-stat Pr(>|t|)
## Easter [1]      0.20370    0.04035   5.048 6.79e-07 ***
## LS (4-2020)    -2.74012    0.10686 -25.642 < 2e-16 ***
## LS (11-2021)   2.08917    0.13192  15.837 < 2e-16 ***
## TC (3-2020)    -0.97239    0.12486  -7.788 5.97e-14 ***
## TC (12-2021)  -1.10141    0.15526  -7.094 6.00e-12 ***
## TC (7-2009)   -0.75267    0.12039  -6.252 1.04e-09 ***
## AO (3-2000)    0.79742    0.14345   5.559 4.99e-08 ***
## TC (12-2020)  -0.63415    0.12443  -5.096 5.36e-07 ***
## LS (3-2003)    0.41613    0.10248   4.061 5.89e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##
## Residual standard error: 0.1617 on 385 degrees of freedom
## Log likelihood = 152.6, aic = 7546, aicc = 7547, bic(corrected for length) = -3.464

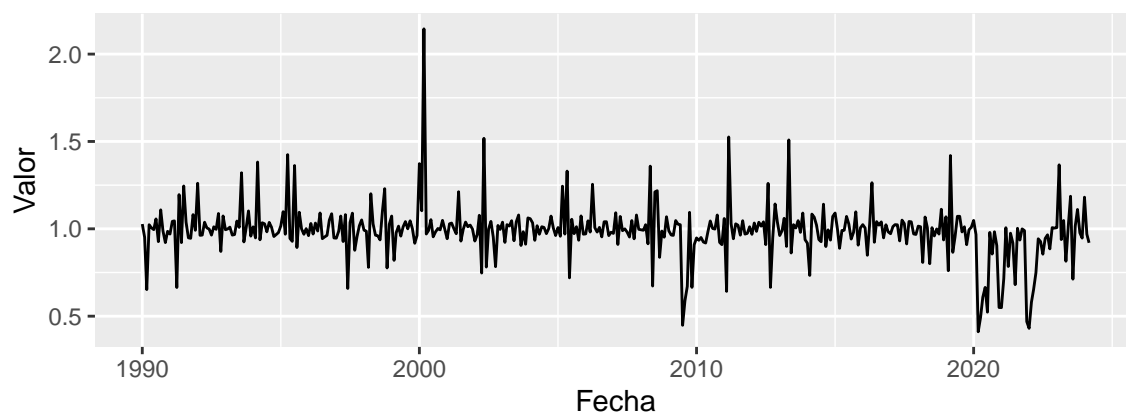
```



Componente estacional



Componente irregular



Predicciones

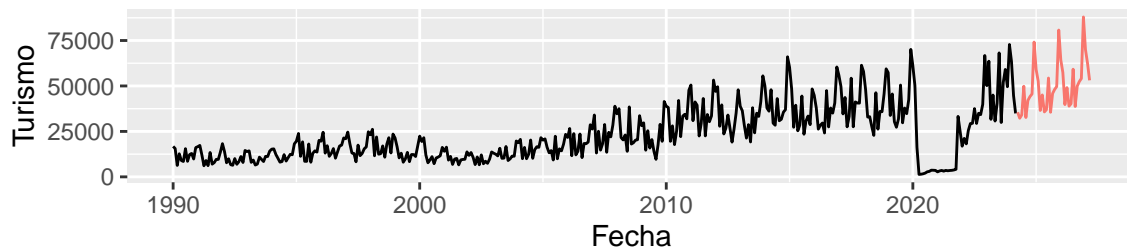
##		y_f	sa_f	t_f	s_f	i_f
##	Apr 2024	35563.42	45558.71	41824.62	0.7806065	1.0892797
##	May 2024	32247.96	41277.68	42182.73	0.7812444	0.9785446
##	Jun 2024	34105.52	42245.74	42544.88	0.8073127	0.9929687
##	Jul 2024	49831.01	44259.54	42898.20	1.1258819	1.0317342
##	Aug 2024	32651.01	42329.02	43283.51	0.7713623	0.9779480
##	Sep 2024	42071.30	44032.25	43785.15	0.9554656	1.0056436
##	Oct 2024	44160.11	43888.13	44648.51	1.0061971	0.9829697
##	Nov 2024	45596.34	45706.42	45814.28	0.9975917	0.9976455
##	Dec 2024	74142.08	46715.70	46947.36	1.5870915	0.9950654
##	Jan 2025	59033.57	48484.93	47837.22	1.2175654	1.0135399
##	Feb 2025	52682.17	48796.00	48271.46	1.0796411	1.0108665
##	Mar 2025	36519.60	49707.54	48196.81	0.7346894	1.0313449
##	Apr 2025	45060.85	46525.80	47815.95	0.9685131	0.9730184
##	May 2025	35720.21	46848.08	47327.48	0.7624690	0.9898705
##	Jun 2025	36945.49	46337.24	46976.16	0.7973174	0.9863991
##	Jul 2025	54367.74	47956.01	46914.59	1.1337004	1.0221982
##	Aug 2025	35541.60	46914.48	47210.49	0.7575828	0.9937299
##	Sep 2025	45829.57	47900.71	47814.52	0.9567618	1.0018026
##	Oct 2025	48093.43	47828.00	48719.40	1.0055497	0.9817034
##	Nov 2025	49661.29	49797.30	49791.12	0.9972685	1.0001242

```

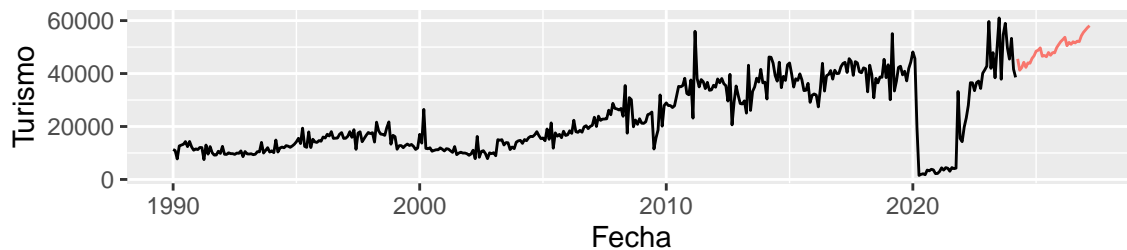
## Dec 2025 80749.82 50985.38 50899.24 1.5837837 1.0016924
## Jan 2026 64295.24 52076.01 51783.47 1.2346423 1.0056493
## Feb 2026 57377.57 52835.55 52277.89 1.0859652 1.0106672
## Mar 2026 39774.49 53705.32 52324.91 0.7406062 1.0263815
## Apr 2026 49076.97 50529.01 52085.97 0.9712633 0.9701078
## May 2026 38903.82 51777.32 51730.51 0.7513680 1.0009048
## Jun 2026 40238.30 51038.47 51458.19 0.7883917 0.9918434
## Jul 2026 59213.33 52014.36 51406.69 1.1384035 1.0118209
## Aug 2026 38709.29 51566.34 51661.63 0.7506696 0.9981556
## Sep 2026 49914.18 52252.48 52224.63 0.9552500 1.0005333
## Oct 2026 52379.80 52057.57 53055.85 1.0061898 0.9811843
## Nov 2026 54087.40 54256.64 54070.05 0.9968807 1.0034509
## Dec 2026 87946.72 55541.51 55178.70 1.5834414 1.0065752
## Jan 2027 70025.62 56450.15 56246.73 1.2404859 1.0036167
## Feb 2027 62491.40 57363.70 57199.31 1.0893893 1.0028739
## Mar 2027 53106.62 58073.72 57962.16 0.9144690 1.0019247

```

Serie original



Serie desestacionalizada



Tendencia

