

Taller 2 Monitoría: Modelación de la Tendencia Determinística

1. Problema

Los datos que se muestran en la Tabla 1 y Figura 1 pertenecen al índice de productividad mensual en Canadá, desde Enero de 1950 hasta Diciembre de 1973, es decir con $N=288$ datos, y están guardados en archivo IndProductividad.txt.

Tabla 1: Índice de productividad mensual en Canadá

| Año | ene | feb | mar | abr | may | jun | jul | ago | sep | oct | nov | dic |
|------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
| 1950 | 77.5 | 77.6 | 78.1 | 78.3 | 78.3 | 78.9 | 79.5 | 80.0 | 80.7 | 82.0 | 82.4 | 82.5 |
| 1951 | 83.4 | 84.4 | 85.8 | 86.5 | 86.8 | 88.0 | 88.7 | 89.4 | 90.2 | 90.6 | 91.3 | 91.4 |
| 1952 | 91.5 | 91.0 | 90.5 | 90.4 | 89.7 | 89.8 | 89.9 | 89.8 | 89.9 | 89.8 | 89.9 | 89.6 |
| 1953 | 89.6 | 89.4 | 88.9 | 88.7 | 88.5 | 88.9 | 89.3 | 89.6 | 89.9 | 90.3 | 89.9 | 89.6 |
| 1954 | 89.6 | 89.6 | 89.4 | 89.5 | 89.4 | 89.9 | 89.9 | 90.6 | 90.4 | 90.4 | 90.4 | 90.2 |
| 1955 | 90.1 | 90.0 | 89.8 | 89.9 | 90.1 | 89.7 | 89.8 | 90.1 | 90.4 | 90.5 | 90.5 | 90.5 |
| 1956 | 90.4 | 90.1 | 90.1 | 90.2 | 90.2 | 91.2 | 91.7 | 92.2 | 92.1 | 92.7 | 93.1 | 93.2 |
| 1957 | 93.1 | 93.3 | 93.3 | 93.6 | 93.7 | 94.1 | 94.3 | 94.9 | 95.4 | 95.5 | 95.4 | 95.3 |
| 1958 | 95.5 | 95.7 | 96.2 | 96.9 | 96.8 | 96.8 | 96.5 | 96.9 | 97.2 | 97.5 | 97.8 | 97.7 |
| 1959 | 97.6 | 97.3 | 97.1 | 97.1 | 97.2 | 96.7 | 97.4 | 97.8 | 98.4 | 99.1 | 99.3 | 99.0 |
| 1960 | 98.7 | 98.5 | 98.2 | 98.7 | 98.6 | 98.8 | 98.7 | 99.0 | 99.4 | 100.2 | 100.3 | 100.3 |
| 1961 | 100.0 | 99.8 | 99.9 | 99.9 | 99.8 | 99.8 | 99.8 | 99.9 | 99.9 | 100.0 | 100.4 | 100.5 |
| 1962 | 100.4 | 100.5 | 100.4 | 100.9 | 100.7 | 101.0 | 101.4 | 101.7 | 101.4 | 101.8 | 102.1 | 102.1 |
| 1963 | 102.2 | 102.2 | 102.2 | 102.4 | 102.4 | 102.8 | 103.3 | 103.6 | 103.3 | 103.4 | 103.7 | 103.9 |
| 1964 | 103.9 | 104.1 | 104.2 | 104.5 | 104.5 | 104.7 | 105.4 | 105.3 | 105.0 | 105.0 | 105.2 | 105.9 |
| 1965 | 106.0 | 106.2 | 106.3 | 106.6 | 106.8 | 107.6 | 108.0 | 107.9 | 107.7 | 107.8 | 108.5 | 109.0 |
| 1966 | 109.3 | 110.0 | 110.2 | 110.8 | 111.0 | 111.3 | 111.7 | 112.2 | 112.3 | 112.5 | 112.6 | 112.9 |
| 1967 | 113.0 | 113.1 | 113.4 | 114.4 | 114.6 | 115.2 | 116.3 | 116.8 | 116.6 | 116.5 | 116.9 | 117.5 |
| 1968 | 118.1 | 118.2 | 118.6 | 119.3 | 119.3 | 119.7 | 120.4 | 120.7 | 121.1 | 121.4 | 121.9 | 122.3 |
| 1969 | 122.6 | 122.6 | 123.2 | 124.6 | 124.9 | 125.9 | 126.4 | 126.9 | 126.6 | 126.8 | 127.4 | 127.9 |
| 1970 | 128.2 | 128.7 | 128.9 | 129.7 | 129.6 | 129.9 | 130.5 | 130.5 | 130.2 | 130.3 | 130.3 | 129.8 |
| 1971 | 130.3 | 130.9 | 131.3 | 132.2 | 132.7 | 133.0 | 134.1 | 135.0 | 134.7 | 134.9 | 135.4 | 136.3 |
| 1972 | 136.7 | 137.3 | 137.4 | 138.2 | 138.3 | 138.5 | 140.2 | 141.3 | 141.8 | 142.0 | 142.3 | 143.3 |
| 1973 | 144.5 | 145.3 | 145.7 | 147.3 | 148.4 | 149.7 | 151.0 | 153.0 | 153.9 | 154.3 | 155.5 | 156.4 |

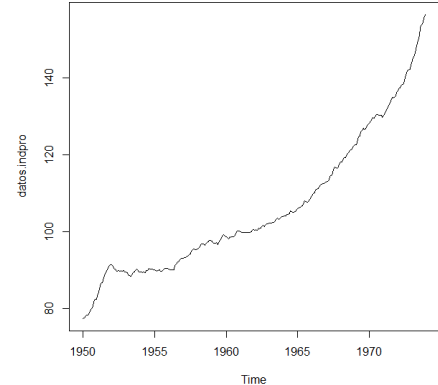


Figura 1. Serie observada

Del análisis gráfico de la serie, responda

1. ¿Qué tipo de tendencia tiene esta serie?
2. ¿Existe componente estacional?
3. ¿Es la serie de componentes aditivas o multiplicativas?
4. Se proponen cuatro modelos aditivos de la forma $Y_t = T_t + E_t$, $E_t \sim IID N(0, \sigma^2)$, y un modelo log polinomial (por tanto, multiplicativo en la escala sin transformar) $\log(Y_t) = T_t^* + E_t$, $E_t \sim IID N(0, \sigma^2)$, para ser ajustados con validación cruzada excluyendo los últimos $m=12$ datos, lo que implica el ajuste con los primeros $n=276$ datos y pronósticos ex – post con los últimos $m=12$ datos:
 - Modelo 1: modelo cúbico
 - Modelo 2: modelo polinomial de grado $p=4$
 - Modelo 3: modelo polinomial de grado $p=6$
 - Modelo 4: modelo exponencial polinomial de grado $p=6$
 - Modelo 5: modelo log polinomial de grado $p=6$

En cada caso, escriba la ecuación del modelo y del modelo ajustado. Concluya si estos modelos son apropiados a partir de los gráficos de ajustes, del análisis de residuales y sus medidas de ajuste y de pronósticos. Evalúe también los gráficos de residuales en términos de la validez del supuesto de independencia, media cero y varianza constante, además de verificar si no hay carencia de ajuste en la representación de la tendencia ¿será apropiado evaluar el supuesto de normalidad en cada modelo? ¿por qué?

5. Considere ahora la serie sin sus primeros cinco años, es decir, desde enero de 1955, y repita 4, de modo que en el ajuste $n=216$ datos.
6. ¿Qué se concluye de 4) y 5) en términos del ajuste, pronósticos y validez de supuestos de los modelos y comportamiento de la serie? ¿las observaciones de los primeros cinco años afectan significativamente la modelación de la serie? ¿por qué?

2. Código R

El siguiente código está disponible en el archivo “Programa R Taller 2 V03.R”, el conjunto de datos se encuentra en el archivo “IndProductividad.txt”.

Parte I: Con los datos desde enero de 1950

```
library(forecast)
#-----
#Cargando funciones de usuario necesarias: Ajuste ruta de archivos a la ruta donde descargue los archivos con las funciones de
#usuario de nombres "Funciones-Criterios.Informacion-Calidad.Intervalos.R" y "Funcion-regexponencial.R"
#-----
#Cargando archivo con las funciones exp.crit.inf.resid() para calcular AIC y BIC, y amplitud.cobertura() para calcular amplitud media y coberturas
#de los I.P
source("C:\\Users\\Nelfi_Gonzalez\\Documents\\ESTADISTICA III 3009137\\CLASES2022\\Talleres de clase\\Funciones de usuario\\Funciones-
Criterios.Informacion-Calidad.Intervalos.R")

#Cargando archivo con la función regexponencial() para hacer la regresión exponencial polinomial
source("C:\\Users\\Nelfi_Gonzalez\\Documents\\ESTADISTICA III 3009137\\CLASES2022\\Talleres de clase\\Funciones de usuario\\Funcion-
regexponencial.R")

#-----
#o bien, desde repositorio en github
#-----
source("https://raw.githubusercontent.com/NelfiGonzalez/Funciones-de-Usuario-Estadistica-III/main/Funciones-Criterios.Informacion-
Calidad.Intervalos.R")
source("https://raw.githubusercontent.com/NelfiGonzalez/Funciones-de-Usuario-Estadistica-III/main/Funcion-regexponencial.R")

#La siguiente línea permite leer los datos guardados en archivo de texto en disco local y
#crear un objeto serie de tiempo. file.choose() abre ventana de windows para
#explorar y ubicar el archivo. header=T para indicar que los datos
#están encabezados por nombre de la columna; si no hay encabezado, se toma header=F
datos.indpro=ts(scan(file.choose(),dec = "."),frequency=12,start=c(1950,1))
datos.indpro

#Gráfica de la serie
win.graph()
plot(datos.indpro)

#Gráfica del logaritmo natural de la serie
win.graph()
plot(log(datos.indpro))

win.graph()
#Gráfica de la descomposición aditiva de la serie
plot(decompose(datos.indpro,type="additive"))

#Gráfica de la tendencia filtrada por filtro de descomposición aditiva
win.graph()
plot(decompose(datos.indpro,type="additive")$trend,ylim=c(min(datos.indpro),max(datos.indpro)),lwd=2)

#Gráfica de la tendencia filtrada por filtro de descomposición aditiva del log de la serie
win.graph()
plot(decompose(log(datos.indpro),type="additive")$trend,ylim=c(min(log(datos.indpro)),max(log(datos.indpro))),lwd=2)

#Boxplots de la distribución de los datos según periodos del año
win.graph(width=5,height=4,points=8)
boxplot(datos.indpro~cycle(datos.indpro),names=month.abb)

#-----
#Definiendo variables y objetos R necesarios en los ajustes y pronósticos
#-----
#Creando índice t de tiempo para el ajuste con los primeros n=276 datos
n=length(datos.indpro)-12 #Tamaño muestra de ajuste
t=1:n
#Potencias del índice de tiempo en el ajuste
t2=t^2
t3=t^3
t4=t^4
t5=t^5
t6=t^6

#Definiendo las matrices de diseño para los ajustes
X1=data.frame(t,t2,t3) #para modelo 1
X2=data.frame(t,t2,t3,t4) #para modelos 2
X3=data.frame(t,t2,t3,t4,t5,t6) #para modelos 3, 4 y 5

yt=ts(datos.indpro[t],freq=12,start=c(1950,1)) #Valores de la serie para el ajuste

tnuevo=(n+1):length(datos.indpro) #índice de tiempo para pronósticos en la validación cruzada
#Potencias del índice de tiempo en los pronósticos
t2nuevo=tnuevo^2
t3nuevo=tnuevo^3
t4nuevo=tnuevo^4
t5nuevo=tnuevo^5
t6nuevo=tnuevo^6

#Definiendo valores de las matrices de diseño en los pronósticos
X1nuevo=data.frame(t=tnuevo,t2=t2nuevo,t3=t3nuevo) #para modelo 1
X2nuevo=data.frame(t=tnuevo,t2=t2nuevo,t3=t3nuevo,t4=t4nuevo) #para modelo 2
X3nuevo=data.frame(t=tnuevo,t2=t2nuevo,t3=t3nuevo,t4=t4nuevo,t5=t5nuevo,t6=t6nuevo) #para modelos 3, 4 y 5

ytf=ts(datos.indpro[tnuevo],freq=12,start=c(1973,1)) #valores de la serie en la validación cruzada
```

```

#-----
#Ajustes y pronósticos de los cinco modelos
#-----
#Ajuste modelo de tendencia cúbica
modelo1=lm(yt~.,data=X1)
summary(modelo1)
ythat1=ts(fitted(modelo1),frequency=12,start=start(yt))
predict1=predict(modelo1,newdata=X1nuevo,interval="prediction") #pronóstico puntual y por I.P modelo 1
predict1=ts(predict1,frequency=12,start=start(ytf))
predict1
#serie de tiempo de los pronósticos puntuales
ytpron1=predict1[,1]

#-----
#Ajuste modelo de tendencia polinomial grado p=4
modelo2=lm(yt~.,data=X2)
summary(modelo2)
ythat2=ts(fitted(modelo2),frequency=12,start=start(yt))

predict2=predict(modelo2,newdata=X2nuevo,interval="prediction") #pronóstico puntual y por I.P modelo 2
predict2=ts(predict2,frequency=12,start=start(ytf))
predict2

#serie de tiempo de los pronósticos puntuales
ytpron2=predict2[,2]

#-----
#Ajuste modelo de tendencia polinomial grado p=6
modelo3=lm(yt~.,data=X3)
summary(modelo3)
ythat3=ts(fitted(modelo3),frequency=12,start=start(yt))

predict3=predict(modelo3,newdata=X3nuevo,interval="prediction") #pronóstico puntual y por I.P modelo 3
predict3=ts(predict3,frequency=12,start=start(ytf))
predict3

#serie de tiempo de los pronósticos puntuales
ytpron3=predict3[,3]

#-----
#Ajuste y pronósticos modelo exponencial polinomial de grado p=6
#Crear vector con nombres de los parámetros a usar en la fórmula R del modelo exponencial
#Forma larga de hacerlo
parammod4=c("beta0","beta1","beta2","beta3","beta4","beta5","beta6")
parammod4
#Forma corta de hacerlo con la ayuda de la función paste0()
parammod4=paste0("beta",0:6)
parammod4
modelo4=regexponencial(respuesta=yt,data=X3,names.param=parammod4) #Ajuste del modelo
summary(modelo4)
ythat4=ts(fitted(modelo4),frequency=12,start=start(yt))

predict4=predict(modelo4,newdata=X3nuevo,interval="prediction") #solo da pronóstico puntual
predict4=ts(predict4,frequency=12,start=start(ytf))
predict4

#serie de tiempo de los pronósticos puntuales
ytpron4=predict4

#-----
#Ajuste modelo 5 log polinomial de grado p=6
modelo5=lm(log(yt)~.,data=X3)
summary(modelo5)
#Por ser modelo log polinomial es necesario destransformar y aplicar factor de corrección exp(MSE/2)
ythat5=ts(exp(fitted(modelo5))*exp(summary(modelo5)$sigma^2/2),freq=12,start=start(yt))

#Por ser modelo log polinomial es necesario destransformar y aplicar factor de corrección
predict5=exp(predict(modelo5,newdata=X3nuevo,interval="prediction"))*exp(summary(modelo5)$sigma^2/2) #pronóstico puntual y por I.P modelo 5
predict5=ts(predict5,frequency=12,start=start(ytf))
predict5

#serie de tiempo de los pronósticos puntuales
ytpron5=predict5[,1]

#-----
#Comparación de los ajustes
#-----
#Gráficos de los ajustes en una misma ventana gráfica
win.graph(width=15,height=10)
layout(rbind(c(1:3),c(4:6)))
plot(datos.indpro,lwd=2)
lines(ythat1,col=2,lwd=2)
legend("topleft",legend=c("Serie real","Serie Ajustada modelo 1"),col=c(1,2),lty=1,lwd=2)
plot(datos.indpro,lwd=2)
lines(ythat2,col=2,lwd=2)
legend("topleft",legend=c("Serie real","Serie Ajustada modelo 2"),col=c(1,2),lty=1,lwd=2)
plot(datos.indpro,lwd=2)
lines(ythat3,col=2,lwd=2)
legend("topleft",legend=c("Serie real","Serie Ajustada modelo 3"),col=c(1,2),lty=1,lwd=2)
plot(datos.indpro,lwd=2)
lines(ythat4,col=2,lwd=2)
legend("topleft",legend=c("Serie real","Serie Ajustada modelo 4"),col=c(1,2),lty=1,lwd=2)
plot(datos.indpro,lwd=2)
lines(ythat5,col=2,lwd=2)
legend("topleft",legend=c("Real","Serie Ajustada modelo 5"),col=c(1,2),lty=1,lwd=2)

#Cálculo AIC y BIC versión exp(Cn*(p)):
#cálculo número de parámetros en cada modelo

```

```

nparmod1=length(coef(modelo1)[coef(modelo1)!=0]);nparmod1
nparmod2=length(coef(modelo2)[coef(modelo2)!=0]);nparmod2
nparmod3=length(coef(modelo3)[coef(modelo3)!=0]);nparmod3
nparmod4=length(coef(modelo4)[coef(modelo4)!=0]);nparmod4
nparmod5=length(coef(modelo5)[coef(modelo5)!=0]);nparmod5
#NOTA: En modelos con ajuste sin transformar los datos, residuals(modelo) calcula Yt-Ythat, pero no en el modelo 5
Criterios1=exp.crit.inf.resid(residuales=residuals(modelo1),n.par=nparmod1);Criterios1
Criterios2=exp.crit.inf.resid(residuales=residuals(modelo2),n.par=nparmod2);Criterios2
Criterios3=exp.crit.inf.resid(residuales=residuals(modelo3),n.par=nparmod3);Criterios3
Criterios4=exp.crit.inf.resid(residuales=residuals(modelo4),n.par=nparmod4);Criterios4
#En el modelo 5, debido a transformación log se usan los pseudo-residuos en lugar de los residuos de ajuste
seudores5=yt-ythat5 #Pseudo-residuos modelo5
Criterios5=exp.crit.inf.resid(residuales=seudores5,n.par=nparmod5);Criterios5

#Organizando en una tabla los valores de AIC y BIC
criterios=rbind(Criterios1,Criterios2,Criterios3,Criterios4,Criterios5)
rownames(criterios) = c("modelo1", "modelo2", "modelo3", "modelo4", "modelo5")
criterios

#-----
#Comparación de residuales de todos los modelos
#-----
#Gráfico de residuales comunes versus tiempo, en una misma ventana gráfica
win.graph(width=15,height=10)
layout(rbind(c(1:3),c(4:6)))
plot.ts(residuals(modelo1),ylim=c(min(residuals(modelo1),-2*summary(modelo1)$sigma,2*summary(modelo1)$sigma),max(residuals(modelo1),-
2*summary(modelo1)$sigma,2*summary(modelo1)$sigma)))
abline(h=c(-2*summary(modelo1)$sigma,0,2*summary(modelo1)$sigma),col=2)
legend("bottomright",legend=c("Modelo 1"),lty=1,col=1,lwd=2)

plot.ts(residuals(modelo2),ylim=c(min(residuals(modelo2),-2*summary(modelo2)$sigma,2*summary(modelo2)$sigma),max(residuals(modelo2),-
2*summary(modelo2)$sigma,2*summary(modelo2)$sigma)))
abline(h=c(-2*summary(modelo2)$sigma,0,2*summary(modelo2)$sigma),col=2)
legend("bottomright",legend=c("Modelo 2"),lty=1,col=1,lwd=2)

plot.ts(residuals(modelo3),ylim=c(min(residuals(modelo3),-2*summary(modelo3)$sigma,2*summary(modelo3)$sigma),max(residuals(modelo3),-
2*summary(modelo3)$sigma,2*summary(modelo3)$sigma)))
abline(h=c(-2*summary(modelo3)$sigma,0,2*summary(modelo3)$sigma),col=2)
legend("topright",legend=c("Modelo 3"),lty=1,col=1,lwd=2)

plot.ts(residuals(modelo4),ylim=c(min(residuals(modelo4),-2*summary(modelo4)$sigma,2*summary(modelo4)$sigma),max(residuals(modelo4),-
2*summary(modelo4)$sigma,2*summary(modelo4)$sigma)))
abline(h=c(-2*summary(modelo4)$sigma,0,2*summary(modelo4)$sigma),col=2)
legend("topright",legend=c("Modelo 4"),lty=1,col=1,lwd=2)

plot.ts(residuals(modelo5),ylim=c(min(residuals(modelo5),-2*summary(modelo5)$sigma,2*summary(modelo5)$sigma),max(residuals(modelo5),-
2*summary(modelo5)$sigma,2*summary(modelo5)$sigma)))
abline(h=c(-2*summary(modelo5)$sigma,0,2*summary(modelo5)$sigma),col=2)
legend("topright",legend=c("Modelo 5"),lty=1,col=1,lwd=2)

#Gráfico de residuales comunes versus variable respuesta, en una misma ventana gráfica
win.graph(width=15,height=10)
layout(rbind(c(1:3),c(4:6)))
plot(fitted(modelo1),residuals(modelo1),cex=1.1,ylim=c(min(residuals(modelo1),-
2*summary(modelo1)$sigma,2*summary(modelo1)$sigma),max(residuals(modelo1),-2*summary(modelo1)$sigma,2*summary(modelo1)$sigma)))
abline(h=c(-2*summary(modelo1)$sigma,0,2*summary(modelo1)$sigma),col=2)
legend("bottomright",legend=c("Modelo 1"),lty=1,col=1,lwd=2)

plot(fitted(modelo2),residuals(modelo2),cex=1.1,ylim=c(min(residuals(modelo2),-
2*summary(modelo2)$sigma,2*summary(modelo2)$sigma),max(residuals(modelo2),-2*summary(modelo2)$sigma,2*summary(modelo2)$sigma)))
abline(h=c(-2*summary(modelo2)$sigma,0,2*summary(modelo2)$sigma),col=2)
legend("bottomright",legend=c("Modelo 2"),lty=1,col=1,lwd=2)

plot(fitted(modelo3),residuals(modelo3),cex=1.1,ylim=c(min(residuals(modelo3),-
2*summary(modelo3)$sigma,2*summary(modelo3)$sigma),max(residuals(modelo3),-2*summary(modelo3)$sigma,2*summary(modelo3)$sigma)))
abline(h=c(-2*summary(modelo3)$sigma,0,2*summary(modelo3)$sigma),col=2)
legend("topright",legend=c("Modelo 3"),lty=1,col=1,lwd=2)

plot(fitted(modelo4),residuals(modelo4),cex=1.1,ylim=c(min(residuals(modelo4),-
2*summary(modelo4)$sigma,2*summary(modelo4)$sigma),max(residuals(modelo4),-2*summary(modelo4)$sigma,2*summary(modelo4)$sigma)))
abline(h=c(-2*summary(modelo4)$sigma,0,2*summary(modelo4)$sigma),col=2)
legend("topright",legend=c("Modelo 4"),lty=1,col=1,lwd=2)

plot(fitted(modelo5),residuals(modelo5),cex=1.1,ylim=c(min(residuals(modelo5),-
2*summary(modelo5)$sigma,2*summary(modelo5)$sigma),max(residuals(modelo5),-2*summary(modelo5)$sigma,2*summary(modelo5)$sigma)))
abline(h=c(-2*summary(modelo5)$sigma,0,2*summary(modelo5)$sigma),col=2)
legend("topright",legend=c("Modelo 5"),lty=1,col=1,lwd=2)

#-----
#Comparación de los pronósticos
#-----
#Gráfico comparativo de pronósticos de modelos 1, 2, 3, 4, 5
plot(ytf,type="b",col=1,pch=19,ylim=c(min(ytf,ytprn1,ytprn2,ytprn3,ytprn4,ytprn5),max(ytf,ytprn1,ytprn2,ytprn3,ytprn4,ytprn5))),lwd=2,ylab="
Ventas en dólares nominales",xaxt="n")
axis(1,at=time(ytf),labels=c("73.1","73.2","73.3","73.4","73.5","73.6","73.7","73.8","73.9","73.10","73.11","73.12"),cex.axis=0.8)
lines(ytprn1,col=2,pch=1,type="b",lwd=2)
lines(ytprn2,col=3,pch=2,type="b",lwd=2)
lines(ytprn3,col=4,pch=3,type="b",lwd=2)
lines(ytprn4,col=5,pch=4,type="b",lwd=2)
lines(ytprn5,col=6,pch=5,type="b",lwd=2)

legend("topleft",legend=c("Real","Modelo 1","Modelo 2","Modelo 3","Modelo 4","Modelo 5"),pch=c(19,1:5),col=c(1:6),lwd=2)

#precisión de los pronósticos por I.P del 95% de confianza
amplcob1=amplitud.cobertura(real=ytf,LIP=predict1[,2],LSP=predict1[,3]);amplcob1
amplcob2=amplitud.cobertura(real=ytf,LIP=predict2[,2],LSP=predict2[,3]);amplcob2
amplcob3=amplitud.cobertura(real=ytf,LIP=predict3[,2],LSP=predict3[,3]);amplcob3
amplcob5=amplitud.cobertura(real=ytf,LIP=predict5[,2],LSP=predict5[,3]);amplcob5

```

```
#Tabla de medidas de precisión de pronósticos
comp=rbind(accuracy(ytpron1,ytf),accuracy(ytpron2,ytf),accuracy(ytpron3,ytf),accuracy(ytpron4,ytf),accuracy(ytpron5,ytf))[,c(2,3,5)]
otros=rbind(amplcob1,amplcob2,amplcob3,c(NA,NA),amplcob5)
tablaprecis=cbind(comp,otros)
rownames(tablaprecis)=c("Modelo 1","Modelo 2","Modelo 3","Modelo 4","Modelo 5")
tablaprecis
```

Parte 2: Ajustes usando sólo los datos desde enero de 1955: La única diferencia con respecto al código usado en la parte I es en el paso de lectura de los datos, lo demás es exactamente igual.

```
rm(list=ls(all=TRUE))
```

```
#-----
#Cargando funciones de usuario necesarias: Ajuste ruta de archivos a la ruta donde descargue los archivos con las funciones de
#usuario de nombres "Funciones-Criterios.Informacion-Calidad.Intervalos.R" y "Funcion-regexponencial.R"
#-----
#Cargando archivo con las funciones exp.crit.inf.resid() para calcular AIC y BIC, y amplitud.cobertura() para calcular amplitud media y coberturas
#de los I.P
source("C:\\Users\\Nelfi_Gonzalez\\Documents\\ESTADISTICA III 3009137\\CLASES2022\\Talleres de clase\\Funciones de usuario\\Funciones-
Criterios.Informacion-Calidad.Intervalos.R")

#Cargando archivo con la función regexponencial() para hacer la regresión exponencial polinomial
source("C:\\Users\\Nelfi_Gonzalez\\Documents\\ESTADISTICA III 3009137\\CLASES2022\\Talleres de clase\\Funciones de usuario\\Funcion-
regexponencial.R")

#-----
#o bien, desde repositorio en github
#-----
source("https://raw.githubusercontent.com/NelfiGonzalez/Funciones-de-Usuario-Estadistica-III/main/Funciones-Criterios.Informacion-
Calidad.Intervalos.R")
source("https://raw.githubusercontent.com/NelfiGonzalez/Funciones-de-Usuario-Estadistica-III/main/Funcion-regexponencial.R")

#-----
#Leyendo datos del archivo IndProductividad.txt excluyendo los cinco primeros años y haciendo gráficos descriptivos usados en este taller
#-----
datos.indpro=ts(scan(file.choose(),dec = ".")[-c(1:60)],frequency=12,start=c(1955,1))
datos.indpro
#Gráfica de la serie
win.graph()
plot(datos.indpro)

#Gráfica del logaritmo natural de la serie
win.graph()
plot(log(datos.indpro))

#Gráfica de la descomposición aditiva de la serie
win.graph()
plot(decompose(datos.indpro,type="additive"))

#Gráfica de la tendencia filtrada por filtro de descomposición aditiva de la serie
win.graph()
plot(decompose(datos.indpro,type="additive")$trend,ylim=c(min(datos.indpro),max(datos.indpro)),lwd=2)

#Gráfica de la tendencia filtrada por filtro de descomposición aditiva del log de la serie
win.graph()
plot(decompose(log(datos.indpro),type="additive")$trend,ylim=c(min(log(datos.indpro)),max(log(datos.indpro))),lwd=2)

#Boxplots de la distribución de los datos según periodos del año
win.graph(width=5,height=4,points=8)
boxplot(datos.indpro~cycle(datos.indpro),names=month.abb)

#-----
#Definiendo variables y objetos R necesarios en los ajustes y pronósticos
#-----
#Creando índice t de tiempo para el ajuste con los primeros 276 datos
n=length(datos.indpro)-12
t=1:n

#Potencias del índice de tiempo en el ajuste
t2=t^2
t3=t^3
t4=t^4
t5=t^5
t6=t^6

#Definiendo las matrices de diseño para los ajustes
X1=data.frame(t,t2,t3) #para modelo 1
X2=data.frame(t,t2,t3,t4) #para modelos 2
X3=data.frame(t,t2,t3,t4,t5,t6) #para modelos 3, 4 y 5

yt=ts(datos.indpro[t],freq=12,start=c(1955,1)) #Valores de la serie para el ajuste

tnuevo=(n+1):length(datos.indpro) #índice de tiempo para pronósticos en la validación cruzada

#Potencias del índice de tiempo en los pronósticos
t2nuevo=tnuevo^2
t3nuevo=tnuevo^3
t4nuevo=tnuevo^4
t5nuevo=tnuevo^5
t6nuevo=tnuevo^6

#Definiendo valores de las matrices de diseño en los pronósticos
X1nuevo=data.frame(t=tnuevo,t2=t2nuevo,t3=t3nuevo) #para modelo 1
X2nuevo=data.frame(t=tnuevo,t2=t2nuevo,t3=t3nuevo,t4=t4nuevo) #para modelo 2
X3nuevo=data.frame(t=tnuevo,t2=t2nuevo,t3=t3nuevo,t4=t4nuevo,t5=t5nuevo,t6=t6nuevo) #para modelos 3, 4 y 5

ytf=ts(datos.indpro[tnuevo],freq=12,start=c(1973,1)) #valores de la serie en la validación cruzada
```

```

#-----
#Ajustes y pronósticos de los modelos
#-----
#Ajuste modelo de tendencia cúbica
modelo1=lm(yt~.,data=X1)
summary(modelo1)
ythat1=ts(fitted(modelo1),frequency=12,start=start(yt))
predict1=predict(modelo1,newdata=X1nuevo,interval="prediction") #pronóstico puntual y por I.P modelo 1
predict1=ts(predict1,frequency=12,start=start(ytf))
predict1
#serie de tiempo de los pronósticos puntuales
ytpron1=predict1[,1]
#-----
#Ajuste modelo de tendencia polinomial grado p=4
modelo2=lm(yt~.,data=X2)
summary(modelo2)
ythat2=ts(fitted(modelo2),frequency=12,start=start(yt))

predict2=predict(modelo2,newdata=X2nuevo,interval="prediction") #pronóstico puntual y por I.P modelo 2
predict2=ts(predict2,frequency=12,start=start(ytf))
predict2

#serie de tiempo de los pronósticos puntuales
ytpron2=predict2[,2]
#-----
#Ajuste modelo de tendencia polinomio p=6
modelo3=lm(yt~.,data=X3)
summary(modelo3)
ythat3=ts(fitted(modelo3),frequency=12,start=start(yt))

predict3=predict(modelo3,newdata=X3nuevo,interval="prediction") #pronóstico puntual y por I.P modelo 3
predict3=ts(predict3,frequency=12,start=start(ytf))
predict3

#serie de tiempo de los pronósticos puntuales
ytpron3=predict3[,3]
#-----
#Ajuste y pronósticos modelo exponencial polinomial de grado p=6
#Crear vector con nombres de los parámetros a usar en la fórmula R del modelo exponencial
#Forma larga de hacerlo
parammod4=c("beta0","beta1","beta2","beta3","beta4","beta5","beta6")
parammod4
#Forma corta de hacerlo con la ayuda de la función paste0()
parammod4=paste0("beta",0:6)
parammod4
modelo4=regexponencial(respuesta=yt,data=X3,names.param=parammod4) #Ajuste del modelo
summary(modelo4)
ythat4=ts(fitted(modelo4),frequency=12,start=start(yt))

predict4=predict(modelo4,newdata=X3nuevo,interval="prediction") #solo da pronóstico puntual
predict4=ts(predict4,frequency=12,start=start(ytf))
predict4

#serie de tiempo de los pronósticos puntuales
ytpron4=predict4
#-----
#Ajuste modelo 5 log polinomial de grado p=6
modelo5=lm(log(yt)~.,data=X3)
summary(modelo5)

#Por ser modelo log polinomial es necesario destransformar y aplicar factor de corrección exp(MSE/2)
ythat5=ts(exp(fitted(modelo5))*exp(summary(modelo5)$sigma^2/2),freq=12,start=start(yt))

#Por ser modelo log polinomial es necesario destransformar y aplicar factor de corrección
predict5=exp(predict(modelo5,newdata=X3nuevo,interval="prediction"))*exp(summary(modelo5)$sigma^2/2) #pronóstico puntual y por I.P modelo 5
predict5=ts(predict5,frequency=12,start=start(ytf))
predict5

#serie de tiempo de los pronósticos puntuales
ytpron5=predict5[,1]
#-----
#Comparación de los ajustes
#-----
#Gráficos de los ajustes en una misma ventana gráfica
win.graph(width=15,height=10)
layout(rbind(c(1:3),c(4:6)))
plot(datos.indpro,lwd=2)
lines(ythat1,col=2,lwd=2)
legend("topleft",legend=c("Serie real","Serie Ajustada modelo 1"),col=c(1,2),lty=1,lwd=2)
plot(datos.indpro,lwd=2)
lines(ythat2,col=2,lwd=2)
legend("topleft",legend=c("Serie real","Serie Ajustada modelo 2"),col=c(1,2),lty=1,lwd=2)
plot(datos.indpro,lwd=2)
lines(ythat3,col=2,lwd=2)
legend("topleft",legend=c("Serie real","Serie Ajustada modelo 3"),col=c(1,2),lty=1,lwd=2)
plot(datos.indpro,lwd=2)
lines(ythat4,col=2,lwd=2)
legend("topleft",legend=c("Serie real","Serie Ajustada modelo 4"),col=c(1,2),lty=1,lwd=2)
plot(datos.indpro,lwd=2)
lines(ythat5,col=2,lwd=2)
legend("topleft",legend=c("Real","Serie Ajustada modelo 5"),col=c(1,2),lty=1,lwd=2)

```

```

#Cálculo AIC y BIC versión exp(Cn*(p))
#Cálculo número de parámetros en cada modelo
nparmod1=length(coef(modelo1)[coef(modelo1)!=0]);nparmod1
nparmod2=length(coef(modelo2)[coef(modelo2)!=0]);nparmod2
nparmod3=length(coef(modelo3)[coef(modelo3)!=0]);nparmod3
nparmod4=length(coef(modelo4)[coef(modelo4)!=0]);nparmod4
nparmod5=length(coef(modelo5)[coef(modelo5)!=0]);nparmod5

#NOTA: En modelos con ajuste sin transformar los datos, residuals(modelo) calcula Yt-Ythat pero no en el modelo 5
Criterios1=exp.crit.inf.resid(residuales=residuals(modelo1),n.par=nparmod1);Criterios1
Criterios2=exp.crit.inf.resid(residuales=residuals(modelo2),n.par=nparmod2);Criterios2
Criterios3=exp.crit.inf.resid(residuales=residuals(modelo3),n.par=nparmod3);Criterios3
Criterios4=exp.crit.inf.resid(residuales=residuals(modelo4),n.par=nparmod4);Criterios4
#En el modelo 5, debido a transformación log se usan los pseudo-residuos en lugar de los residuos de ajuste
seudores5=yt-ythat5 #Seudo-residuos modelo5
Criterios5=exp.crit.inf.resid(residuales=seudores5,n.par=nparmod5);Criterios5
#Organizando en una tabla los valores de AIC y BIC
criterios=rbind(Criterios1,Criterios2,Criterios3,Criterios4,Criterios5)
rownames(criterios) = c("modelo1","modelo2","modelo3","modelo4","modelo5")
criterios

#-----
#Comparación de residuales de todos los modelos
#-----
#Gráfico de residuales comunes versus tiempo, en una misma ventana gráfica
win.graph(width=15,height=10)
layout(rbind(c(1:3),c(4:6)))
plot.ts(residuals(modelo1),ylim=c(min(residuals(modelo1),-2*summary(modelo1)$sigma),max(residuals(modelo1),-
2*summary(modelo1)$sigma),2*summary(modelo1)$sigma)))
abline(h=c(-2*summary(modelo1)$sigma,0,2*summary(modelo1)$sigma),col=2)
legend("bottomright",legend=c("Modelo 1"),lty=1,col=1,lwd=2)

plot.ts(residuals(modelo2),ylim=c(min(residuals(modelo2),-2*summary(modelo2)$sigma),max(residuals(modelo2),-
2*summary(modelo2)$sigma),2*summary(modelo2)$sigma)))
abline(h=c(-2*summary(modelo2)$sigma,0,2*summary(modelo2)$sigma),col=2)
legend("bottomright",legend=c("Modelo 2"),lty=1,col=1,lwd=2)

plot.ts(residuals(modelo3),ylim=c(min(residuals(modelo3),-2*summary(modelo3)$sigma),max(residuals(modelo3),-
2*summary(modelo3)$sigma),2*summary(modelo3)$sigma)))
abline(h=c(-2*summary(modelo3)$sigma,0,2*summary(modelo3)$sigma),col=2)
legend("topright",legend=c("Modelo 3"),lty=1,col=1,lwd=2)

plot.ts(residuals(modelo4),ylim=c(min(residuals(modelo4),-2*summary(modelo4)$sigma),max(residuals(modelo4),-
2*summary(modelo4)$sigma),2*summary(modelo4)$sigma)))
abline(h=c(-2*summary(modelo4)$sigma,0,2*summary(modelo4)$sigma),col=2)
legend("topright",legend=c("Modelo 4"),lty=1,col=1,lwd=2)

plot.ts(residuals(modelo5),ylim=c(min(residuals(modelo5),-2*summary(modelo5)$sigma),max(residuals(modelo5),-
2*summary(modelo5)$sigma),2*summary(modelo5)$sigma)))
abline(h=c(-2*summary(modelo5)$sigma,0,2*summary(modelo5)$sigma),col=2)
legend("topright",legend=c("Modelo 5"),lty=1,col=1,lwd=2)

#Gráfico de residuales comunes versus variable respuesta, en una misma ventana gráfica
win.graph(width=15,height=10)
layout(rbind(c(1:3),c(4:6)))
plot(fitted(modelo1),residuals(modelo1),cex=1.1,ylim=c(min(residuals(modelo1),-
2*summary(modelo1)$sigma),max(residuals(modelo1),-2*summary(modelo1)$sigma),2*summary(modelo1)$sigma)))
abline(h=c(-2*summary(modelo1)$sigma,0,2*summary(modelo1)$sigma),col=2)
legend("bottomright",legend=c("Modelo 1"),lty=1,col=1,lwd=2)

plot(fitted(modelo2),residuals(modelo2),cex=1.1,ylim=c(min(residuals(modelo2),-
2*summary(modelo2)$sigma),max(residuals(modelo2),-2*summary(modelo2)$sigma),2*summary(modelo2)$sigma)))
abline(h=c(-2*summary(modelo2)$sigma,0,2*summary(modelo2)$sigma),col=2)
legend("bottomright",legend=c("Modelo 2"),lty=1,col=1,lwd=2)

plot(fitted(modelo3),residuals(modelo3),cex=1.1,ylim=c(min(residuals(modelo3),-
2*summary(modelo3)$sigma),max(residuals(modelo3),-2*summary(modelo3)$sigma),2*summary(modelo3)$sigma)))
abline(h=c(-2*summary(modelo3)$sigma,0,2*summary(modelo3)$sigma),col=2)
legend("topright",legend=c("Modelo 3"),lty=1,col=1,lwd=2)

plot(fitted(modelo4),residuals(modelo4),cex=1.1,ylim=c(min(residuals(modelo4),-
2*summary(modelo4)$sigma),max(residuals(modelo4),-2*summary(modelo4)$sigma),2*summary(modelo4)$sigma)))
abline(h=c(-2*summary(modelo4)$sigma,0,2*summary(modelo4)$sigma),col=2)
legend("topright",legend=c("Modelo 4"),lty=1,col=1,lwd=2)

plot(fitted(modelo5),residuals(modelo5),cex=1.1,ylim=c(min(residuals(modelo5),-
2*summary(modelo5)$sigma),max(residuals(modelo5),-2*summary(modelo5)$sigma),2*summary(modelo5)$sigma)))
abline(h=c(-2*summary(modelo5)$sigma,0,2*summary(modelo5)$sigma),col=2)
legend("topright",legend=c("Modelo 5"),lty=1,col=1,lwd=2)

#-----
#Comparación de los pronósticos
#-----
#Gráfico comparativo de pronósticos de modelos 1, 2, 3, 4, 5
plot(ytf,type="b",col=1,pch=19,ylim=c(min(ytf,ytpron1,ytpron2,ytpron3,ytpron4,ytpron5),max(ytf,ytpron1,ytpron2,ytpron3,ytpron4,ytpron5))),lwd=2,ylab="
Ventas en dólares nominales",xaxt="n")
axis(1,at=time(ytf),labels=c("73.1","73.2","73.3","73.4","73.5","73.6","73.7","73.8","73.9","73.10","73.11","73.12"),cex.axis=0.8)
lines(ytpron1,col=2,pch=1,type="b",lwd=2)
lines(ytpron2,col=3,pch=2,type="b",lwd=2)
lines(ytpron3,col=4,pch=3,type="b",lwd=2)
lines(ytpron4,col=5,pch=4,type="b",lwd=2)
lines(ytpron5,col=6,pch=5,type="b",lwd=2)

legend("topleft",legend=c("Real","Modelo 1","Modelo 2","Modelo 3","Modelo 4","Modelo 5"),pch=c(19,1:5),col=c(1:6),lwd=2)

```

```

#precisión de los pronósticos por I.P del 95% de confianza
amplcob1=amplitud.cobertura(real=ytf,LIP=predict1[,2],LSP=predict1[,3]);amplcob1
amplcob2=amplitud.cobertura(real=ytf,LIP=predict2[,2],LSP=predict2[,3]);amplcob2
amplcob3=amplitud.cobertura(real=ytf,LIP=predict3[,2],LSP=predict3[,3]);amplcob3
amplcob5=amplitud.cobertura(real=ytf,LIP=predict5[,2],LSP=predict5[,3]);amplcob5
#Tabla de medidas de precisión de pronósticos
comp=rbind(accuracy(ytpron1,ytf),accuracy(ytpron2,ytf),accuracy(ytpron3,ytf),accuracy(ytpron4,ytf),accuracy(ytpron5,ytf))[,c(2,3,5)]
otros=rbind(amplcob1,amplcob2,amplcob3,c(NA,NA),amplcob5)

tablaprecis=cbind(comp,otros)
rownames(tablaprecis)=c("Modelo 1","Modelo 2","Modelo 3","Modelo 4","Modelo 5")
tablaprecis

```