Act8 Series de tiempo no estacionarias C_IACD_Estadística

A01750164 | Paul Martín García Morfín

2022-11-21

Series de tiempo no estacionarias

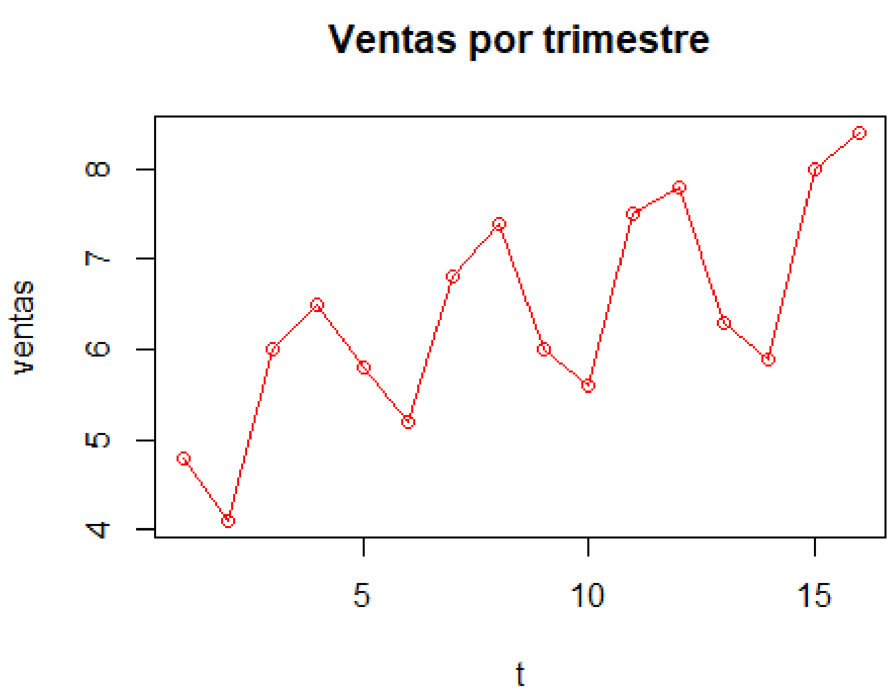
Problema 1: Venta de televisores

Usa los datos de las ventas de televisores para familiarizarte con el análisis de tendencia de una serie de tiempo:

Año 1 2 3 4
Trimestre 1 2 3 4 1 2 3 4 1 2 3 4 1 2 3 4 Ventas (miles) 4.8 4.1 6.0 6.5 5.8 5.2 6.8 7.4 6.0 5.6 7.5 7.8 6.3 5.9 8.0 8.4

1. Realiza el gráfico de dispersión. Observa la tendencia y los ciclos

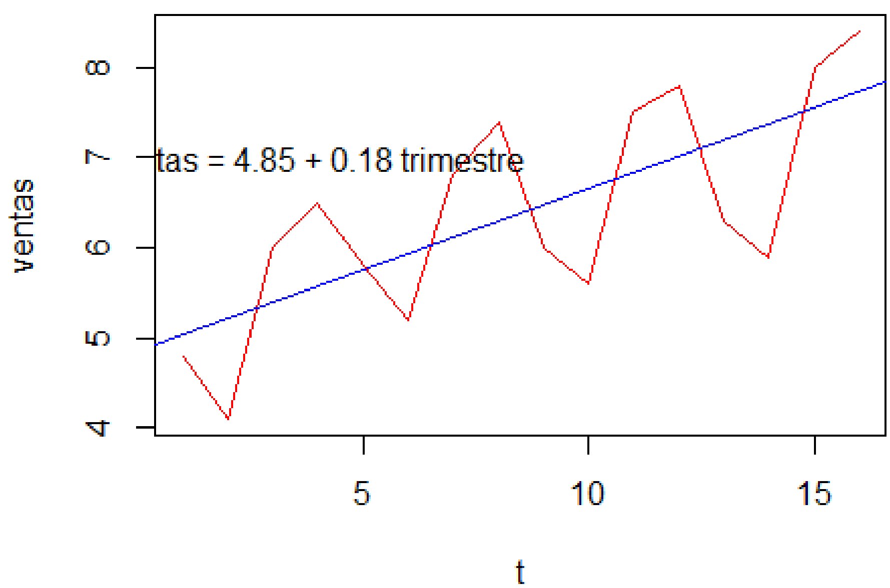
```
t = 1:16
ventas = c(4.8, 4.1, 6.0, 6.5, 5.8, 5.2, 6.8, 7.4, 6.0, 5.6, 7.5, 7.8,
6.3, 5.9, 8.0, 8.4)
plot(t, ventas, type="o", col="red", main="Ventas por trimestre")
```



Graficando los datos proporcionados, se puede observar que corresponden a una serie de tiempo no estacionaria ya que se visualiza un claro crecimiento. Cuando se pone un negocio se espera que el negocio prospere y las ventas crezcan, es normal que haya fluctuaciones en el año pero que la tendencia es que crezca.

2. Realiza el análisis de tendencia y estacionalidad

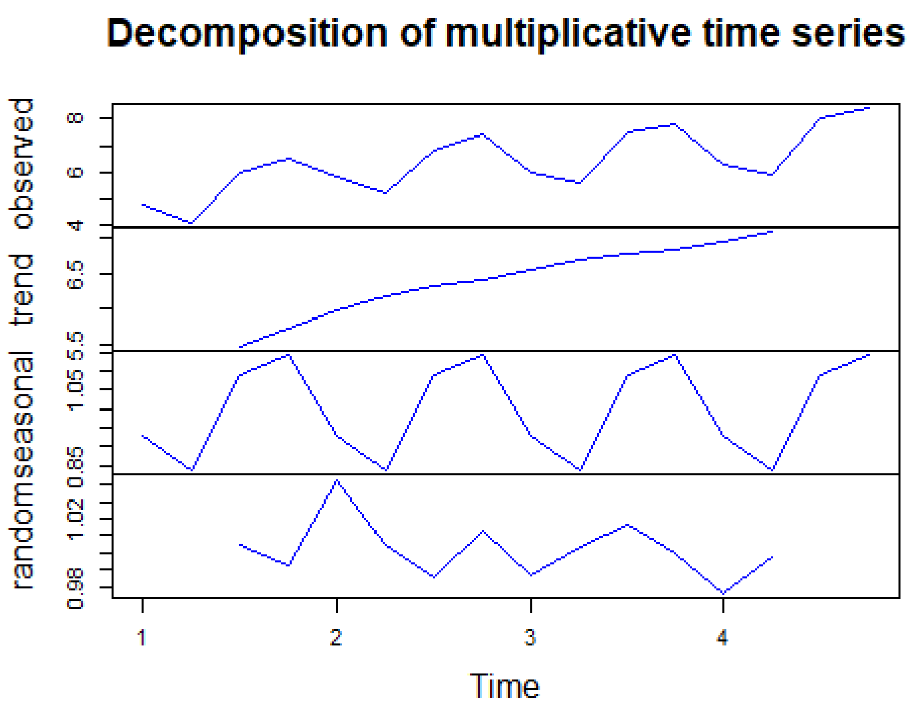
```
N = lm(ventas~t)
plot(t, ventas, type="l", col="red")
abline(N, col = "blue")
N

##
## Call:
## lm(formula = ventas ~ t)
##
## Coefficients:
## (Intercept)            t
##      4.8525       0.1799

text(4, 7, "ventas = 4.85 + 0.18 trimestre")
```



Se puede trazar una línea recta y observar cómo la tendencia va creciendo, sin embargo, el problema con esta tendencia es que no está tomando en cuenta la estacionalidad, que son los ciclos a lo largo del año.

2.1. Descompón la serie en sus 3 componentes e interprétalos

```
ventas = c(4.8, 4.1, 6, 6.5, 5.8, 5.2, 6.8, 7.4, 6, 5.6, 7.5, 7.8, 6.3,
5.9, 8, 8.4)
x = ts(ventas, frequency=4, start(c(2016,1)))
T = decompose(x, type="m")
plot(T, col ="blue")
```



La primer gráfica son los valores originales observados, mientras que las otras tres son los componentes, los cuales se explican a continuación:

- Tendencia: La tendencia de la serie de tiempo se caracteriza por un patrón gradual y consistente de las variaciones de la propia serie. La tendencia a largo plazo se ajusta al esquema de moverse continuamente hacía arriba. Es la línea recta encontrada.

- Variación estacional: Este componente representa la variabilidad en los datos debido a las estaciones, la cual corresponde a los movimientos de la serie que recurren año tras año en los mismos trimestres (para este caso) del año, con la misma intensidad.

- Variación irregular: Se debe a factores a corto plazo que afectan a la serie de tiempo y que son imprevisibles. Son los errores, es decir, la diferencia entre lo predicho y lo esperado.

3. Analiza el modelo lineal de la tendencia

3.1. Realiza la regresión lineal de la tendencia (ventas desestacionalizadas vs tiempo)

```
ventas_desestacionalizadas = (T$x)/(T$seasonal)
x2 = 1:16
y2 = ventas_desestacionalizadas
N2 = lm(y2~x2)
N2

##
## Call:
## lm(formula = y2 ~ x2)
##
## Coefficients:
## (Intercept)           x2
##      5.1080       0.1474
```

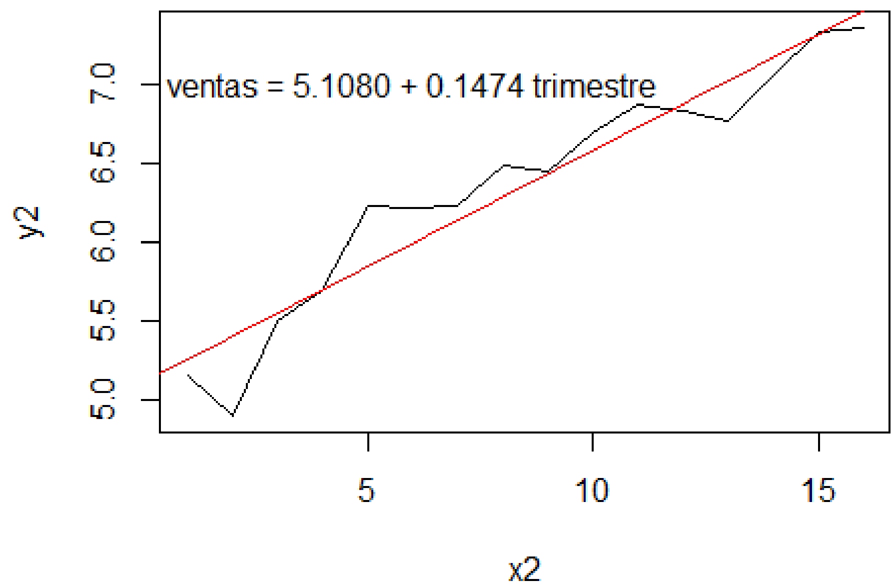
Las ventas desestacionalizadas son las ventas a las que ya se les quitó dos cosas:

- La irregularidad debido al azar, con los promedios móviles.
- La estacionalidad, con los índices estacionales (fluctuación debido a las estaciones del año).

Se aplica regresión lineal simple y se obtiene la recta, cuya es y = 5.1080 + 0.1474x

3.2. Dibuja la recta junto con las ventas desestacionalizadas

```
plot(x2, y2, type="l")
abline(N2, col="red")
text(6, 7, "ventas = 5.1080 + 0.1474 trimestre")
```



Como se puede observar en la gráfica, se logró un suavizamiento de la serie diferente. Sobre esta se trabaja la regresión lineal para encontrar la tendencia. Es importante mencionar que se usó un esquema multiplicativo (donde la tendencia se va incrementando) para este suavizamiento.

3.3. Pertinencia del modelo lineal

Significancia de β1

```
summary(N2)

##
## Call:
## lm(formula = y2 ~ x2)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -0.5007 -0.1001  0.0037  0.1207  0.3872
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  5.10804    0.11171   45.73  < 2e-16 ***
## x2           0.14738    0.01155   12.76 4.25e-09 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.213 on 14 degrees of freedom
```

```
## Multiple R-squared:  0.9208, Adjusted R-squared:  0.9151
## F-statistic: 162.7 on 1 and 14 DF,  p-value: 4.248e-09
```

Se rechaza la hipótesis nula ya que $p < \alpha$ y podemos decir que β_1 es signfificativamente diferente de 0.

Variabilidad explicada por el modelo (c.d)

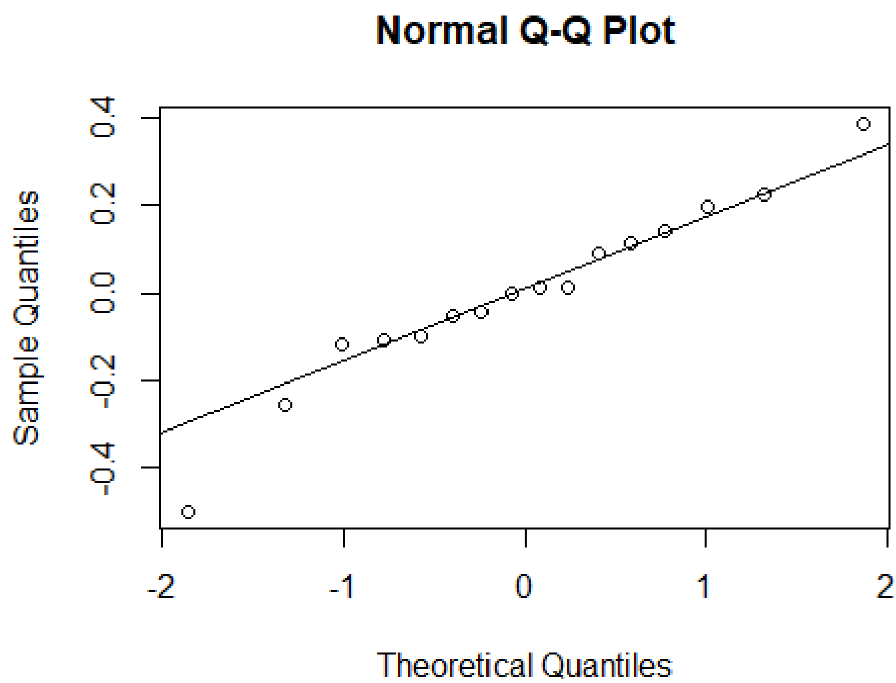
```
cor = cor(x2, y2)
cd = cor^2
cd

## [1] 0.9207911
```

El modelo explica el 92.08% de la variabilidad. Se deduce que aproximadamente el 8% de las ventas no se relaciona con el tiempo.

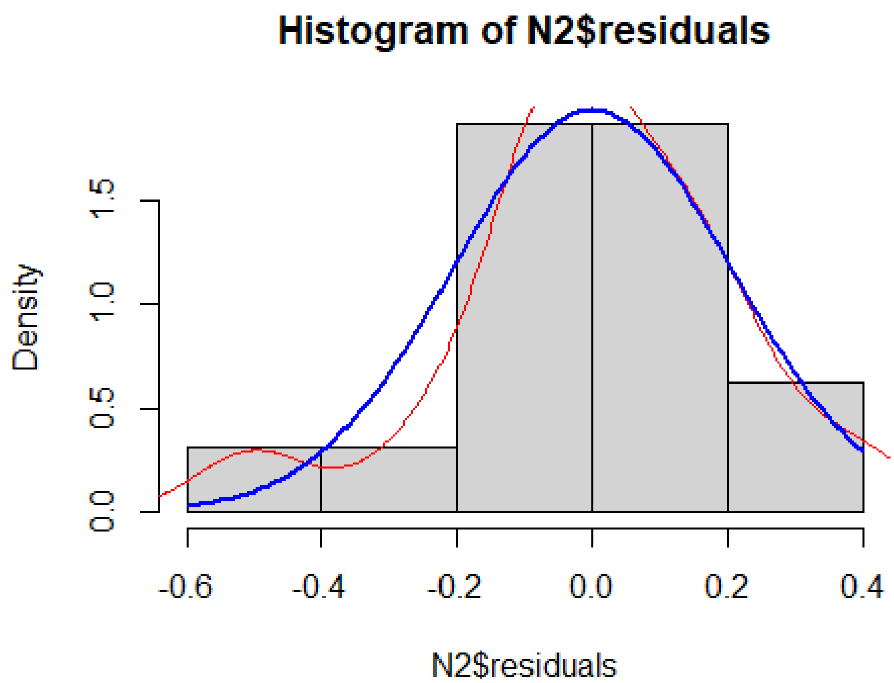
Análisis de los residuos

- Normalidad de los residuos

```
shapiro.test(N2$residuals)

##
##  Shapiro-Wilk normality test
##
## data:  N2$residuals
## W = 0.96379, p-value = 0.7307

qqnorm(N2$residuals)
qqline(N2$residuals)
```




```
hist(N2$residuals, freq=FALSE)
lines(density(N2$residual), col="red")
curve(dnorm(x, mean=mean(N2$residuals), sd=sd(N2$residuals)), add=TRUE,
col="blue", lwd=2)
```



El p-value es mayor a alfa, así que nos quedamos con la hipótesis nula y podemos decir que los datos provienen de una normal, asi que los residuos e_i también se distribuyen como una normal.

- Verificación de media cero

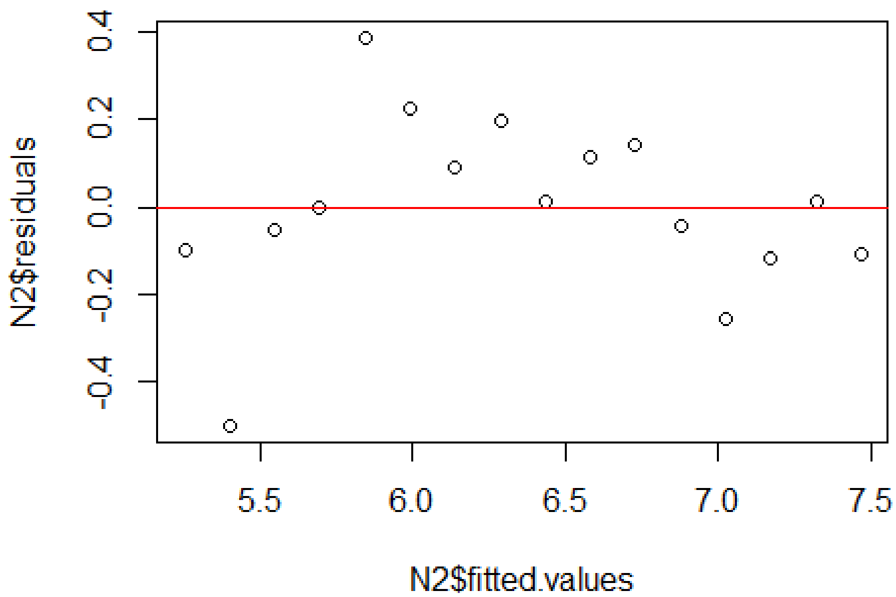
```
t.test(N2$residuals)

##
##  One Sample t-test
##
## data:  N2$residuals
## t = -1.4751e-16, df = 15, p-value = 1
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
##  -0.1096629  0.1096629
## sample estimates:
##     mean of x
## -7.589415e-18
```

El p-value es mayor a alfa, así que nos quedamos con la hipótesis nula y podemos decir que la media de los residuos no es significativamente diferente de cero.

Tenemos que $E(e_i) = 0$

- Homocedasticidad

```
plot(N2$fitted.values, N2$residuals)
abline(h=0, col="red")
```



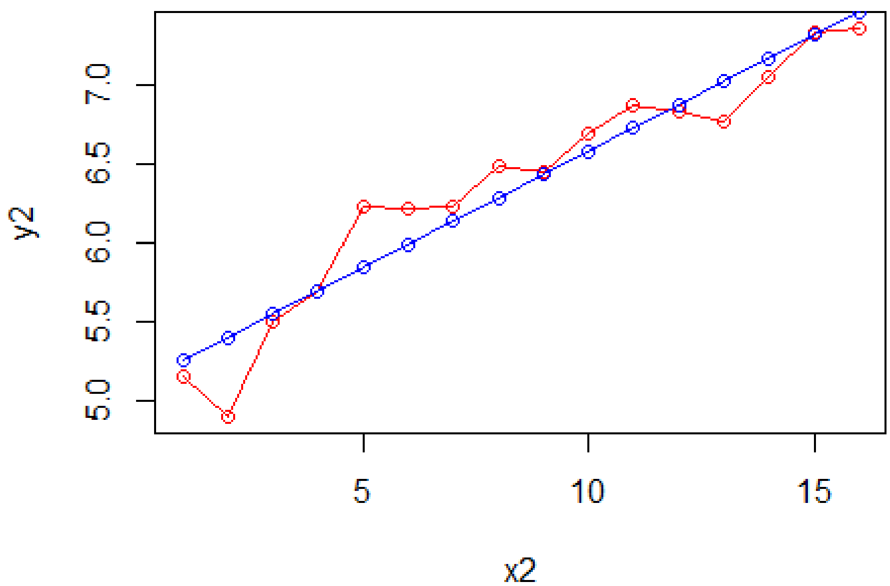
Parece no existir homocedasticidad en los residuos ya que hay heterocedasticidad y sesgo, es decir, presenta asimetría.

4. Calcula el CME y el EPAM (promedio de los errores porcentuales) de la predicción de la serie de tiempo

```
p = NA
e = NA
er = NA

for(i in 1:16){
  p[i] = 5.1080 + 0.1474*i
  e[i] = y2[i] - p[i]
  er[i] = (abs(e[i])/y2[i])*100
}
T2 = data.frame(x2, p, y2, e^2, er)
CME = mean(e^2, na.rm=TRUE)
EPAM = mean(er, na.rm=TRUE)
cat("CME = ", CME)

## CME =  0.03970642

cat("\nEPAM = ", EPAM)

##
## EPAM =  2.439396
```

5. Dibuja el gráfico de los valores de las ventas y las predicciones vs el tiempo

```
plot(x2, y2, type="o", col="red")
lines(x2, p[x2], type="o", col="blue")
```



6. Concluye sobre el modelo

Gracias a las métricas utilizadas y a los análisis realizados se podría concluir que es un buen modelo, ya que se obtuvo un CME pequeño y el coeficiente de determinación indica un alto porcentaje de explicación, además de que en la mayoría de pruebas de validación de residuos se cumplió con los supuestos, sin embargo, en los tests de homocedasticidad no salió como lo esperado, es decir, presentó sesgo y heterocedasticidad en los residuos. Esto se puede deber a que el comportamiento no es del todo lineal, o hace falta considerar otro factor para explicar la variabilidad completa. Sería necesario hacer algunos ajustes o transformaciones para mejorar en ese sentido.

Por otro lado, se remarca la importancia de estas técnicas para el análisis de variables que varían con el tiempo, especialmente cuando se trata de casos parecidos al presentado en este trabajo como lo son el análisis de ventas o problemas de economía en general, ya que pueden ayudar al manejo de riesgos y toma de decisiones. Sin embargo, también es importante validar el modelo y recordar que lo que se obtendrá son sólo pronósticos que pueden no cumplirse debido a factores externos.

7. Realiza el pronóstico para el siguiente año

```
f = function(x) {5.1080 + 0.1474*x}
# Los ídices estacionales son:
```

```r
a1 = T$seasonal[1]
a2 = T$seasonal[2]
a3 = T$seasonal[3]
a4 = T$seasonal[4];
f(17)*a1*1000

## [1] 7085.872

f(18)*a2*1000

## [1] 6491.284

f(19)*a3*1000

## [1] 8632.585

f(20)*a4*1000

## [1] 9195.263
```

Para este cálculo se estableció la función lineal obtenida anteriormente y se sacaron los índices estacionales de la matriz al hacer la descomposición (para así tomar en cuenta la estacionalidad), se multiplica por la función evaluada en el tiempo de cada trimestre y, para obtener las unidades correctas se multiplica por 1000.

Problema 2: Venta de libros

A continuación, se presentan los datos correspondientes a los últimos tres años de ventas trimestrales (número de ejemplares vendidos) de un libro de texto universitario.

Trimestre Año 1 Año 2 Año 3 1 1690 1800 1850 2 940 900 1100 3 2625 2900 2930 4 2500 2360 2615

1. Encuentre los promedios móviles de cuatro trimestres y los promedios móviles centrados

```r
libros = c(1690, 940, 2625, 2500, 1800, 900, 2900, 2360, 1850, 1100,
2930, 2615)
t2 = c(1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12)

pm = NA
pmc = NA

for(i in 1:(12-4)){
  pm[i+4] = (libros[i]+libros[i+1]+libros[i+2]+libros[i+3])/4
}
for(i in 1:(12-6)){
  pmc[i+6] = (pm[i+4]+pm[i+5])/2
}
```

```r
PM = data.frame(t2, libros, pm, pmc)
PM
```

```
##    t2 libros      pm      pmc
## 1   1   1690      NA       NA
## 2   2    940      NA       NA
## 3   3   2625      NA       NA
## 4   4   2500      NA       NA
## 5   5   1800 1938.75       NA
## 6   6    900 1966.25       NA
## 7   7   2900 1956.25 1952.500
## 8   8   2360 2025.00 1961.250
## 9   9   1850 1990.00 1990.625
## 10 10   1100 2002.50 2007.500
## 11 11   2930 2052.50 1996.250
## 12 12   2615 2060.00 2027.500
```

2. Calcule los índices estacionales de los cuatro trimestres

```r
vsi = NA

for(i in 1:(12-6)){
  vsi[i+6] = libros[i+6]/pmc[i+6]
}

VSI = data.frame(t2, libros, pmc, vsi)
is = VSI$vsi[9:12]
is
```

```
## [1] 0.9293564 0.5479452 1.4677520 1.2897657
```

3. ¿Cuándo obtiene la editorial el mayor índice estacional? ¿Parece razonable este resultado? ¿Por qué?

En el tercer trimestre del año.