

Técnicas de Predicción: Regresión y series de tiempo

Evelyn Gutierrez

Predicción

Métodos simples para la predicción.

Métodos simples para la predicción

Dada la serie:

$$y_1, y_2, \dots, y_T$$

Queremos predecir:

$$y_{T+h|T}$$

- Método de la media

$$\hat{y}_{T+h|T} = \bar{y} = (y_1 + \dots + y_T)/T$$

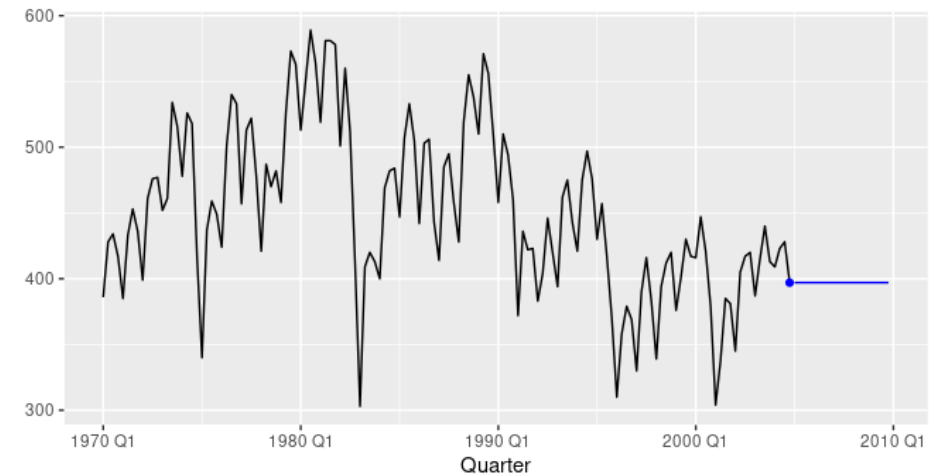
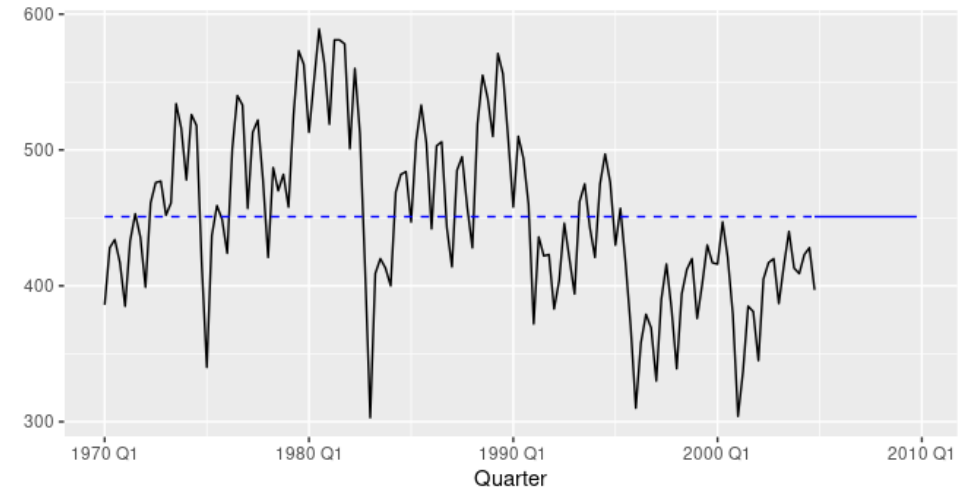
El valor predicho es la media de todo.

- Método Naïve

$$\hat{y}_{T+h|T} = y_T$$

El valor predicho es el último valor de la serie.

También se llama “Random Walk” porque funciona bien en una serie paseo aleatorio.



Métodos simples para la predicción

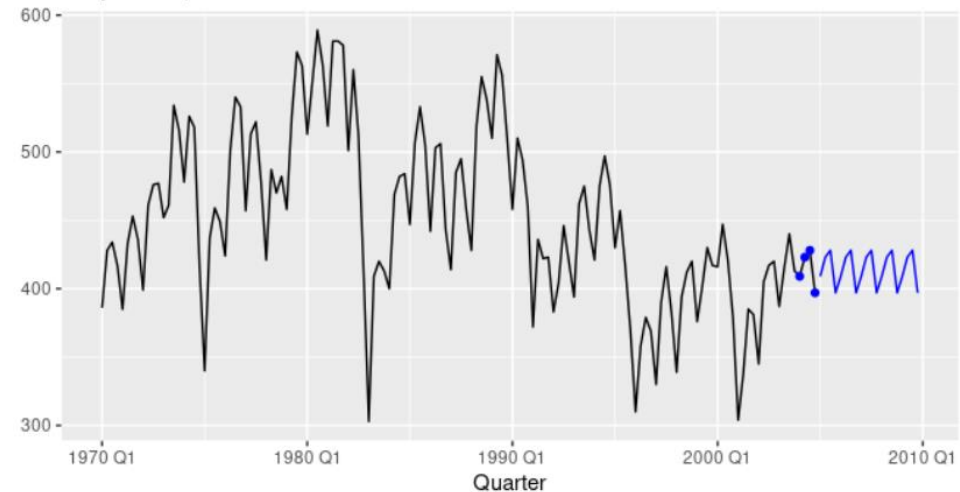
- Método Naïve Estacional

$$\hat{y}_{T+h|T} = y_{T+h-m(k+1)}:$$

m : periodo estacional

$$h = \text{floor}\left(\frac{h-1}{m}\right)$$

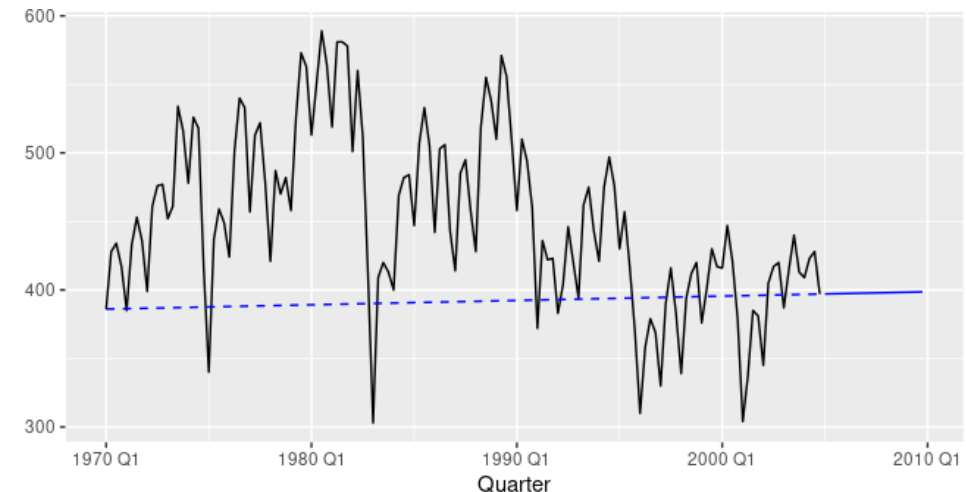
El valor predicho repite el último periodo estacional.



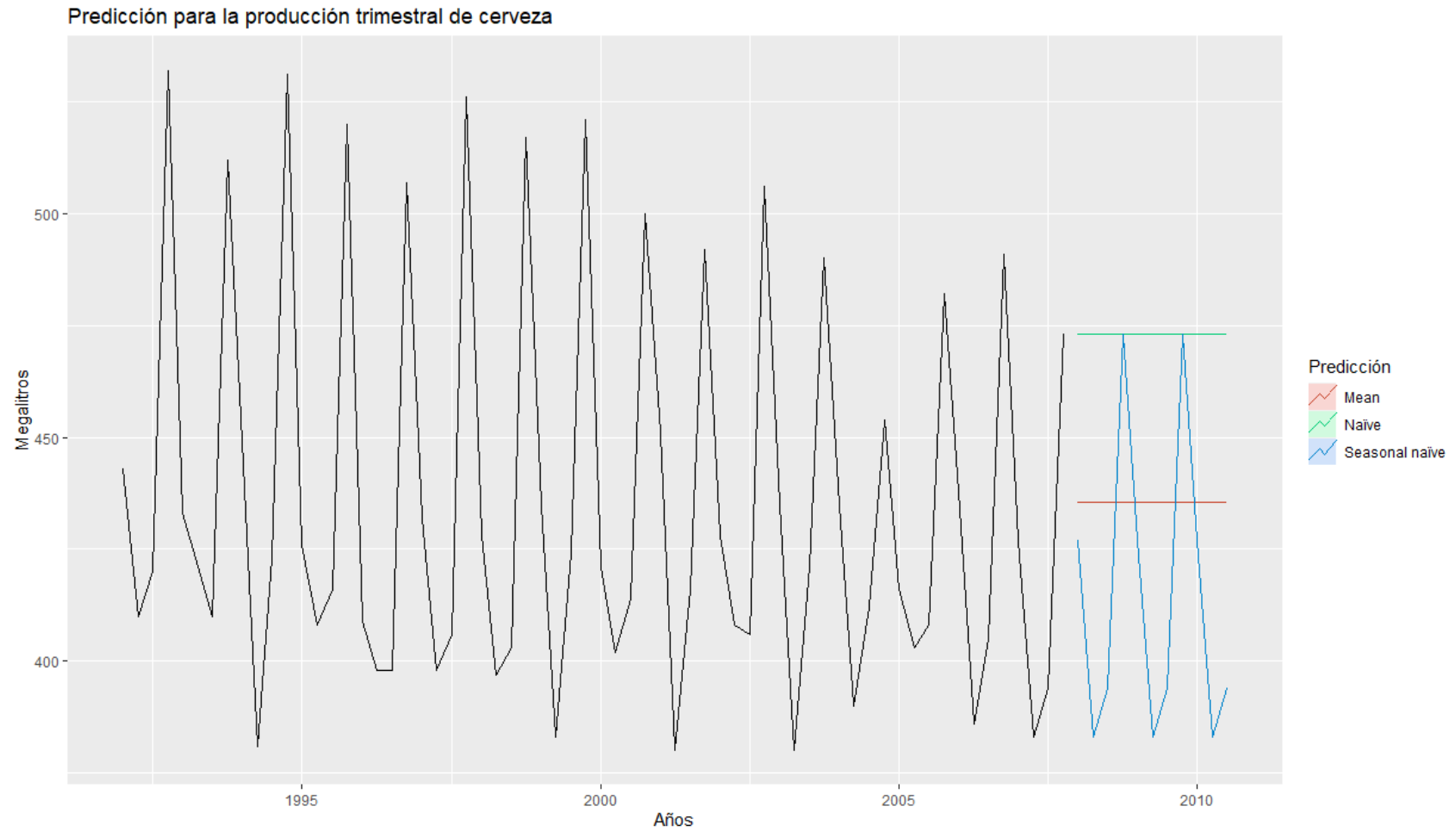
- Método de la deriva (drift)

$$\hat{y}_{T+h|T} = y_T + \frac{h}{T-1} \sum_{t=2}^T (y_t - y_{t-1}) = y_T + h \left(\frac{y_T - y_1}{T-1} \right)$$

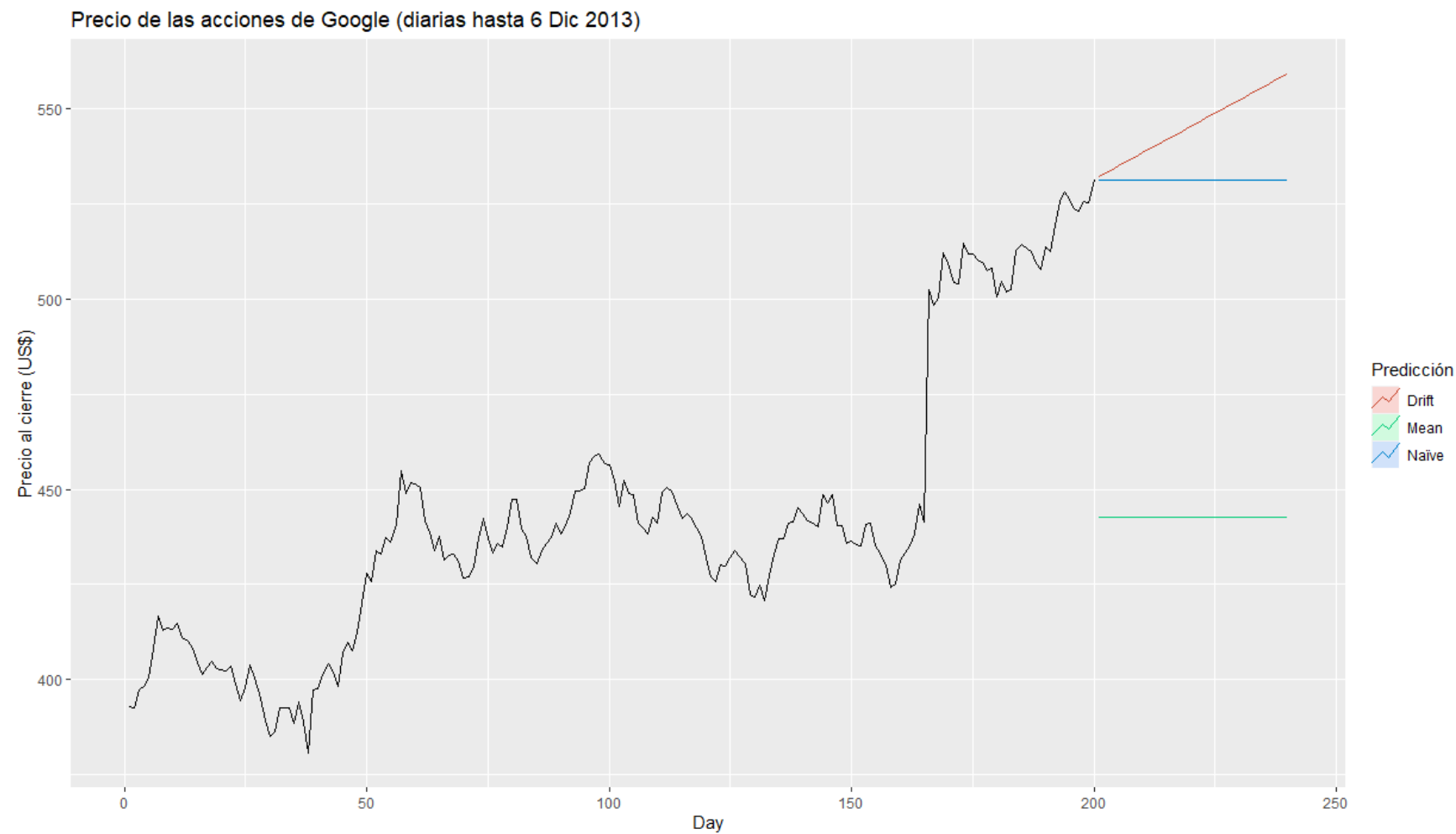
El valor predicho el cambio promedio de la serie.



Ejemplo



Ejemplo



Regresión lineal

En el contexto de series de tiempo

Regresión Lineal

- Supongamos que la serie y_t esta influenciada por variables independientes

$$x_{t1}, x_{t2}, x_{t3}, \dots, x_{tq}$$

- La relación planteada es una relación lineal

$$y_t = \beta_1 x_{t1} + \beta_2 x_{t2} + \dots + \beta_q x_{tq} + w_t$$

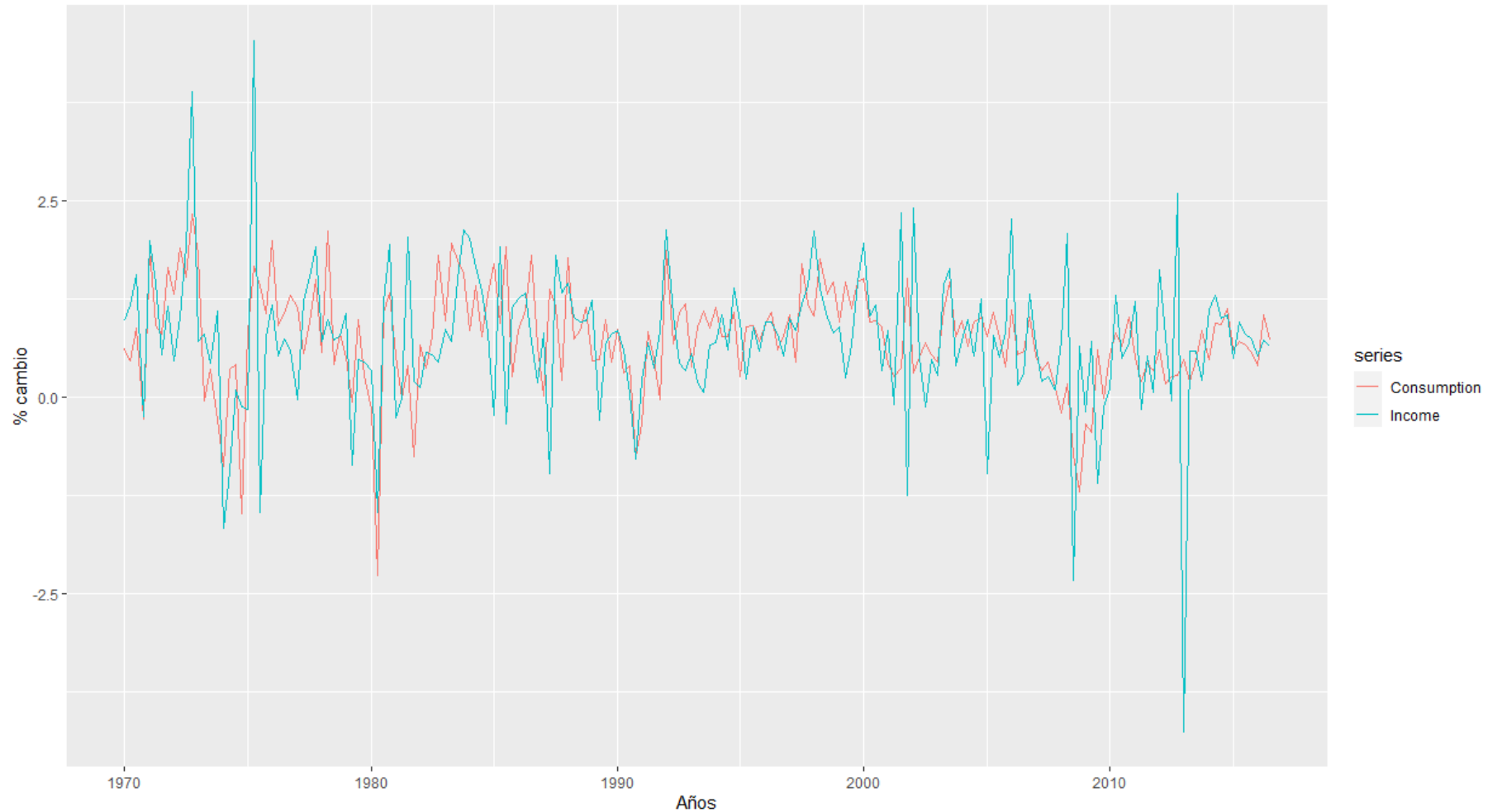
donde

- $\{\beta_i\}_{i=1,\dots,q}$ son coeficientes de regresión
- $\{w_t\}$ error aleatorio o proceso de ruido independiente normal de media cero y varianza σ_w^2

Para una regresión de series de tiempo, el error no suele ser ruido blanco. Relajaremos un poco los supuestos.

Ejemplo

Se tienen los datos trimestrales para la tasa de crecimiento del gasto de consumo personal (y) e ingreso disponible personal (x) en EE. UU. desde el primer trimestre de 1970 hasta el tercer trimestre de 2016.

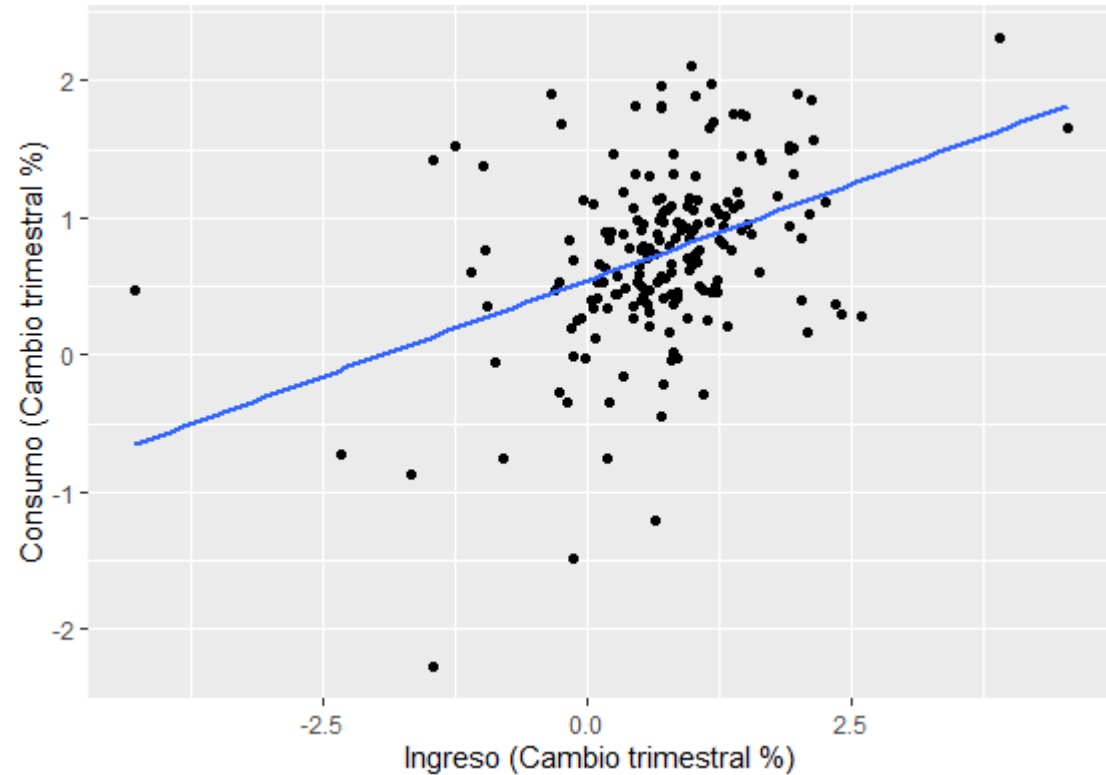


Ejemplo

Se tienen los datos trimestrales para la tasa de crecimiento del gasto de consumo personal (y) e ingreso disponible personal (x) en EE. UU. desde el primer trimestre de 1970 hasta el tercer trimestre de 2016.

Call:
`tslm(formula = Consumption ~ Income, data = uschange)`

Coefficients:
(Intercept) Income
0.5451 0.2806



Regresión lineal múltiple

También será posible el uso de regresiones con múltiples variables regresoras

Las métricas estudiadas para la evaluación de la bondad de ajuste del modelo serán las que se estudiaron previamente (R^2 ajustado)

Una verificación especial que debe realizarse es sobre los residuales.

Call:

```
tslm(formula = Consumption ~ Income + Production + Unemployment + Savings, data = uschange)
```

Residuals:

Min	1Q	Median	3Q	Max
-0.88296	-0.17638	-0.03679	0.15251	1.20553

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)	
(Intercept)	0.26729	0.03721	7.184	1.68e-11	***
Income	0.71449	0.04219	16.934	< 2e-16	***
Production	0.04589	0.02588	1.773	0.0778	.
Unemployment	-0.20477	0.10550	-1.941	0.0538	.
Savings	-0.04527	0.00278	-16.287	< 2e-16	***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.3286 on 182 degrees of freedom
Multiple R-squared: 0.754, Adjusted R-squared: 0.7486
F-statistic: 139.5 on 4 and 182 DF, p-value: < 2.2e-16

Verificación de los supuestos

Serie de tiempo

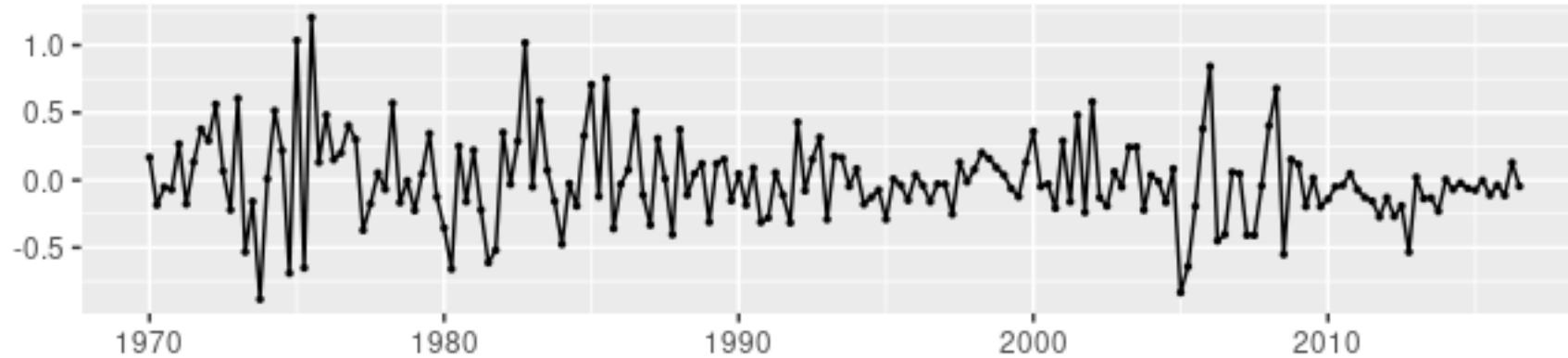
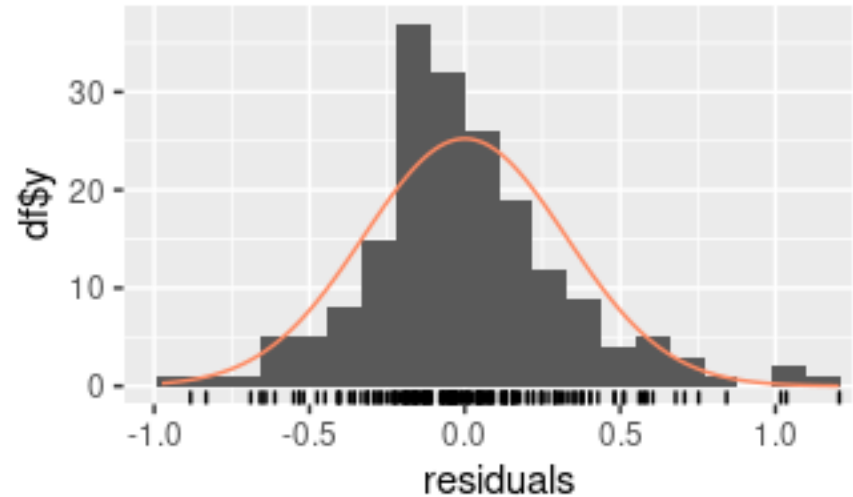
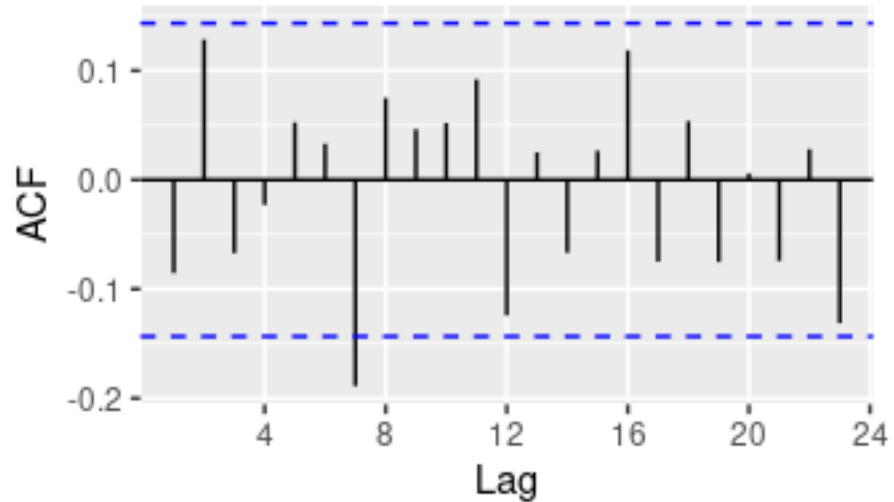


Gráfico de autocorrelación



Hisotgrama de los residuales

Verificación de los supuestos

- Test para la autocorrelación en los residuales
 - Test de **Breusch-Godfrey**
 - Hipótesis nula: No hay correlación hasta el orden p

```
> checkresiduals(fit.consMR)
```

```
Breusch-Godfrey test for serial correlation of order up  
to 8
```

```
data: Residuals from Linear regression model  
LM test = 14.874, df = 8, p-value = 0.06163
```

Predictores útiles en series de tiempo:

- Tendencia

$$y_t = \beta_0 + \beta_1 t + \varepsilon_t$$

- Variables indicadoras de estacionalidad

Podemos usar variable que capturen los patrones estacionales a través del uso de variables indicadoras

	d_1t	d_2t	d_3t	d_4t	d_5t	d_6t
Lunes	1	0	0	0	0	0
Martes	0	1	0	0	0	0
Miercoles	0	0	1	0	0	0
Jueves	0	0	0	1	0	0
Viernes	0	0	0	0	1	0
Sabado	0	0	0	0	0	1
Domingo	0	0	0	0	0	0
Lunes	1	0	0	0	0	0
Martes	0	1	0	0	0	0
Miercoles	0	0	1	0	0	0
Jueves	0	0	0	1	0	0

$$y_t = \beta_0 + \beta_1 d_{1t} + \dots + \beta_6 d_{6t}$$

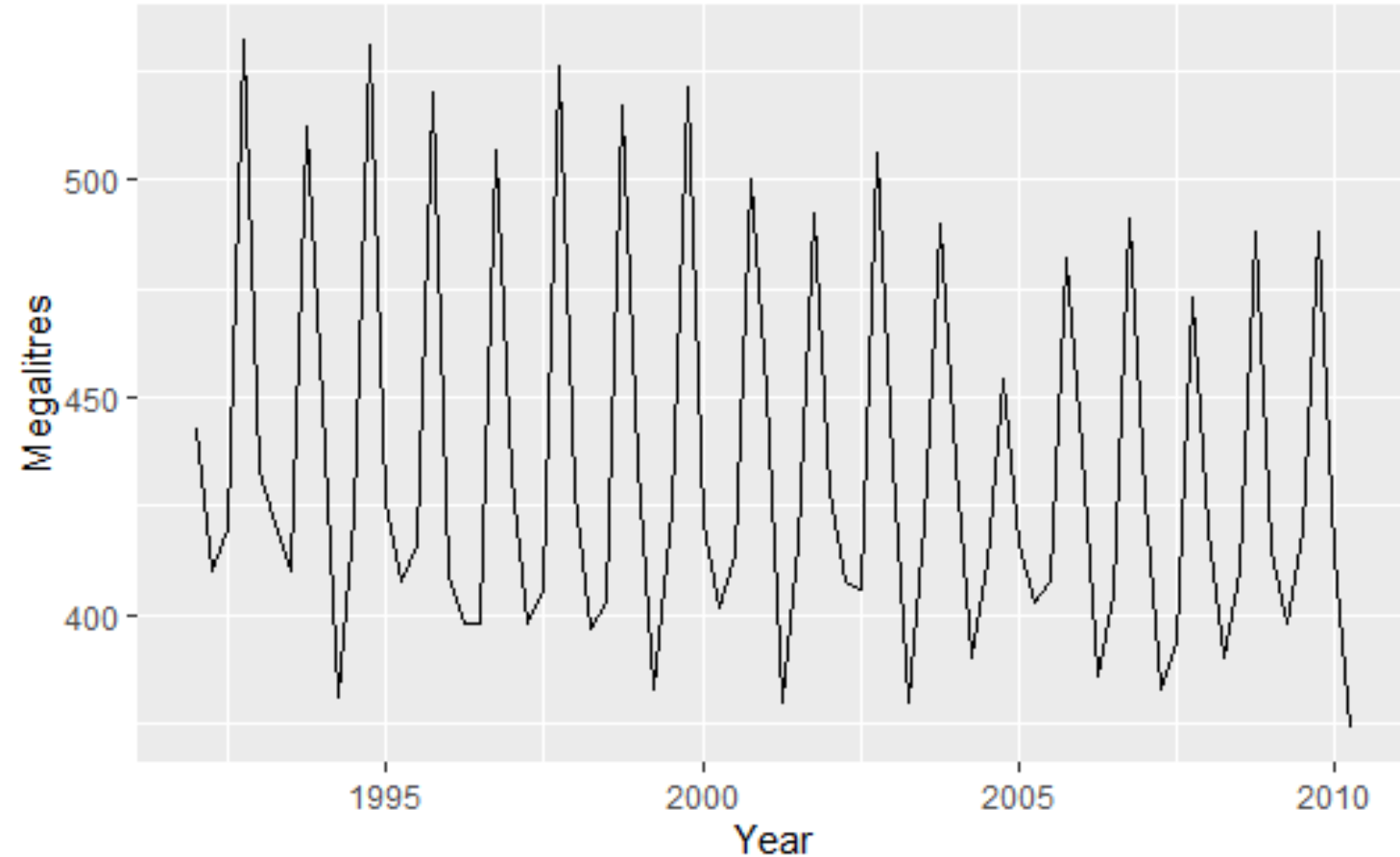
Interpretación del coeficiente para d_{it} :

Efecto del día i relativo al día 7 (domingo)

Ejemplo: Regresión lineal

La serie de la producción de cerveza entre el año 1992 y 2011 se modela utilizando una regresión lineal con:

- Tendencia
- Variables indicadores estacionales



Ejemplo: Regresión lineal

Se ajusta una regresión lineal con la tendencia y variables indicadoras para cada trimestre.

- season1: Trim1 (Referencia)
- season2: Trim2
- season3: Trim3
- season4: Trim4

```
Call:
tslm(formula = beer2 ~ trend + season)

Residuals:
    Min       1Q   Median       3Q      Max
-42.903  -7.599  -0.459   7.991  21.789

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  441.80044    3.73353  118.333  < 2e-16 ***
trend        -0.34027    0.06657   -5.111  2.73e-06 ***
season2     -34.65973    3.96832   -8.734  9.10e-13 ***
season3     -17.82164    4.02249   -4.430  3.45e-05 ***
season4      72.79641    4.02305   18.095  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 12.23 on 69 degrees of freedom
Multiple R-squared:  0.9243,    Adjusted R-squared:  0.9199
F-statistic: 210.7 on 4 and 69 DF,  p-value: < 2.2e-16
```


Ejemplo: Regresión lineal

- Tendencia a la baja de -0,34 megalitros por trimestre.
- En promedio, el 2do trimestre tiene producción de 34.7 megalitros menor que el primer trimestre.
- En promedio, el 3r trimestre tiene una producción de 17.8 megalitros menor que el primer trimestre.
- Finalmente, en promedio, el 4to trimestre tiene una producción de 72.8 megalitros mayor que el primer trimestre.

Call:

```
tslm(formula = beer2 ~ trend + season)
```

Residuals:

Min	1Q	Median	3Q	Max
-42.903	-7.599	-0.459	7.991	21.789

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)	
(Intercept)	441.80044	3.73353	118.333	< 2e-16	***
trend	-0.34027	0.06657	-5.111	2.73e-06	***
season2	-34.65973	3.96832	-8.734	9.10e-13	***
season3	-17.82164	4.02249	-4.430	3.45e-05	***
season4	72.79641	4.02305	18.095	< 2e-16	***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 12.23 on 69 degrees of freedom

Multiple R-squared: 0.9243, Adjusted R-squared: 0.9199

F-statistic: 210.7 on 4 and 69 DF, p-value: < 2.2e-16