



UNIVERSITAT DE  
BARCELONA



# Anàlisi de l'atur registrat a Catalunya en els últims anys

ANÀLISIS DE SÈRIES TEMPORALS

LAURA JULIÀ MELIS

07.11.2018

---

*En aquest informe s'analitza una base de dades amb informació relativa a l'atur a Catalunya (en milers de persones aturades, és a dir, sense feina però disponibles i buscant ocupació) amb l'objectiu d'estudiar quin tipus de sèrie temporal és aquesta i fer prediccions de les darreres observacions amb els diferents mètodes deterministes estudiats a l'assignatura.*

---

# Índex

1.	Introducció.....	3
2.	Aplicació empírica.....	4
2.2.	Tipologia de la sèrie. ....	4
2.2.1.	Contrast de Daniel. ....	4
2.2.2.	Contrast de Kruskal-Wallis.....	5
2.3.	Anàlisi determinista de la sèrie.....	5
2.3.1.	Mètode ingenu. ....	6
2.3.2.	Mètode de les dobles mitjanes mòbils.....	7
2.3.3.	Mètode allisat exponencial de Holt (AEH). ....	8
2.3.4.	Mètode de la tendència lineal.....	9
3.	Conclusions.....	10

## 1. Introducció.

En aquest document es recull una anàlisi determinista sobre l'atur registrat a Catalunya en els últims anys. Per entendre millor l'objecte d'estudi d'aquest informe, cal entendre bé què s'entén per **atur** i com es mesura. L'**atur** d'un país és el nombre de població activa que no treballa, és a dir, una persona es troba en situació d'atur quan no té un lloc de feina però voldria treballar. Les dades es recullen mirant el nombre d'aturats inscrits a les oficines d'ocupació de Catalunya cada mes.

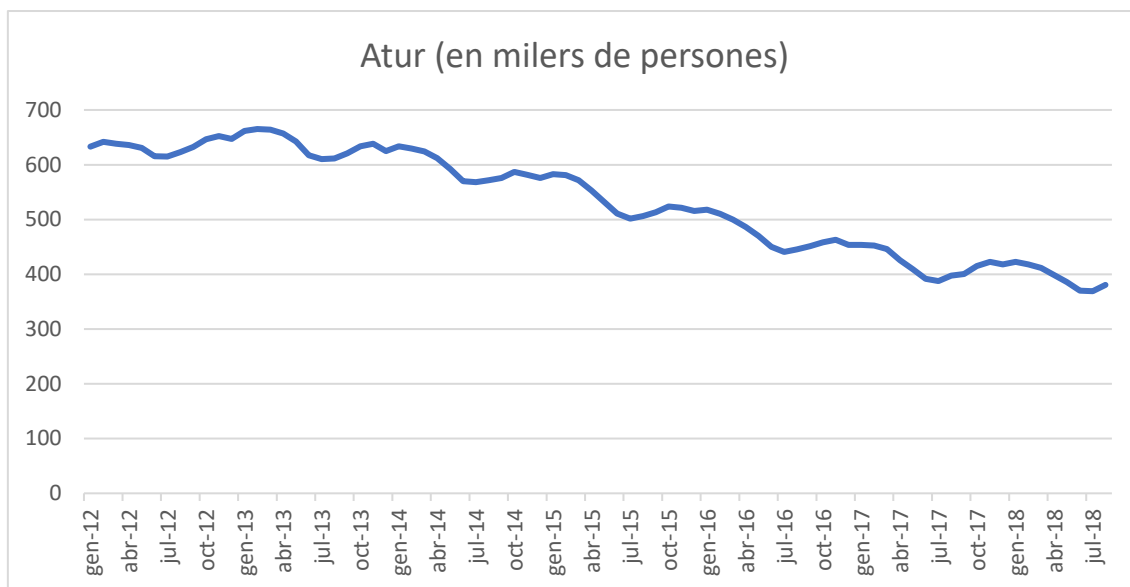
La informació sobre l'atur registrat a Catalunya (xifres en milers de persones) s'ha obtingut de la pàgina web de l'IDESCAT (<https://www.idescat.cat/indicadors/?id=conj&n=10220&col=1>) i la font de les dades és el Departament de Treball, Afers Socials i Famílies. Originalment, la base de dades contenia l'atur segregat per sexe i grups d'edat, així com també l'atur total des del gener de l'any 1996 fins a l'agost del 2018. Per al present estudi, només s'ha utilitzat la xifra d'atur dels darrers 6 anys i 8 mesos (01/2012 – 08/2018) i, com que es tracten de dades mensuals, s'ha obtingut finalment una sèrie de 80 registres. A més, s'han dividit dos períodes en les dades: el període mostral conté les dades compreses entre els anys 2012 i 2017, i l'extra mostral, les de l'any 2018.

L'anàlisi en qüestió es farà en diverses passes. En primer lloc, per tal de conèixer la tipologia de la sèrie temporal (si té o no tendència i/o estacionalitat), es realitzaran dos contrastos d'hipòtesis. A continuació i depenent dels resultats obtinguts, s'aplicaran els mètodes de predicció més adequats (anàlisi determinista) i es calcularà l'error quadràtic mig (EQM), l'error absolut mig (EAM) i l'error percentual absolut mig (EPAM) de cada un per tal de decidir quin és el mètode que ens ofereix les millors prediccions (o millor dit, les prediccions menys dolentes). Per concloure, es farà una petita discussió de resultats i una decisió final.

## 2. Aplicació empírica.

### 2.2. Tipologia de la sèrie.

Primer, per tal de tenir una idea general de com són les nostres dades, s'ha fet el gràfic de línies que es mostra a continuació.



#### 2.2.1. Contrast de Daniel.

Aquest test es realitza amb la finalitat de conèixer si la sèrie temporal té o no tendència. Les hipòtesis són:

$H_0$ : La sèrie no té tendència.

$H_1$ : La sèrie sí que té tendència.

Es tracta de comparar l'estadístic  $Z$  amb el valor crític d'una Normal (0,1) de dues cues en valor absolut. L'estadístic  $Z$  es calcula com:

$$Z = \sqrt{T-1} * Ts \quad \text{amb} \quad Ts = 1 - \frac{6 \sum_{t=1}^T d_t^2}{T(T^2 - 1)}$$

On  $T$  és el nombre d'observacions ( $T=80$ ),  $t$  és la variable de tendència determinista (creada assignant a cada observació un valor de 1 fins a  $T$ ) i  $d_t$ , la diferència entre el rang i  $t$  (el rang és la posició que ocupa cada observació quan ordenem les dades de menor a major).

Un cop realitzats aquets càlculs (mirar pestanya **Test de Daniel** a l'Excel), hem obtingut que  $|Z| = 8.4825$  i que el valor crític amb un nivell de significació del 5% és 1.96. Així doncs, com que  $|Z| > 1.96$ , rebutgem la hipòtesi nul·la i diem que la sèrie té tendència.

A més, mirant la gràfica anterior es pot comprovar com la sèrie té tendència decreixent.

### 2.2.2. Contrast de Kruskal-Wallis.

Amb aquest test es contrasten les següents hipòtesis:

$H_0$ : La sèrie no té component estacional.

$H_1$ : La sèrie sí té component estacional.

I ara es tracta de comparar l'estadístic de Kruskal-Wallis en valor absolut amb el valor crític d'una Chi-quadrat amb  $s-1$  graus de llibertat, on  $s$  és el nombre de períodes estacionals dintre d'un any (en aquest cas,  $s=12$  ja que les dades són mensuals). L'estadístic K-W es calcula com:

$$KW = \frac{12}{T(T-1)} * \left[ \sum_{i=1}^s \frac{(R_i)^2}{T_i} \right] - 3(T+1)$$

On  $T_i$  és el nombre d'observacions que corresponen a l'estació  $i$ -èssima i  $R_i$ , el nombre de rangs corresponents a l'estació  $i$ -èssima. El rang  $i$  i  $T$  són els valors comentats en el contrast anterior.

Després d'haver realitzat els càlculs corresponents (mirar pestanya **Test K-W** a l'Excel) s'ha obtingut que  $KW=6.198$  i que el valor crític amb un nivell de significació del 5% i 11 graus de llibertat és 19.675. Per tant, com que  $|KW| < 19.675$ , no es rebutja la hipòtesi nul·la i direm que la sèrie temporal no té estacionalitat (malgrat que amb la representació gràfica sembla que sí).

En conseqüència, la sèrie objecte d'estudi és una sèrie de tipus III (sèrie amb tendència i sense component estacional) així que els mètodes per a l'anàlisi clàssica que es realitzarà a continuació seran, a més del mètode ingenu, el mètode de les dobles mitjanes mòbils, l'allisat exponencial de Holt i el de tendència lineal.

### 2.3. Anàlisi determinista de la sèrie.

Per a cada un dels mètodes predictius que s'aplicaran, es calcularan els següents errors de predicció (per als dos períodes):

- Error Absolut Mitjà.

$$EAM = \sum_{t=1}^T \frac{|e_t|}{T} = \sum_{t=1}^T \frac{|\hat{y}_t - y_t|}{T}$$

- Error Quadràtic Mig.

$$EQM = \sum_{t=1}^T \frac{e_t^2}{T} = \sum_{t=1}^T \frac{(\hat{y}_t - y_t)^2}{T}$$

- Error Percentual Absolut Mig.

$$EPAM = \frac{100}{T} \sum_{t=1}^T \left| \frac{e_t}{y_t} \right|$$

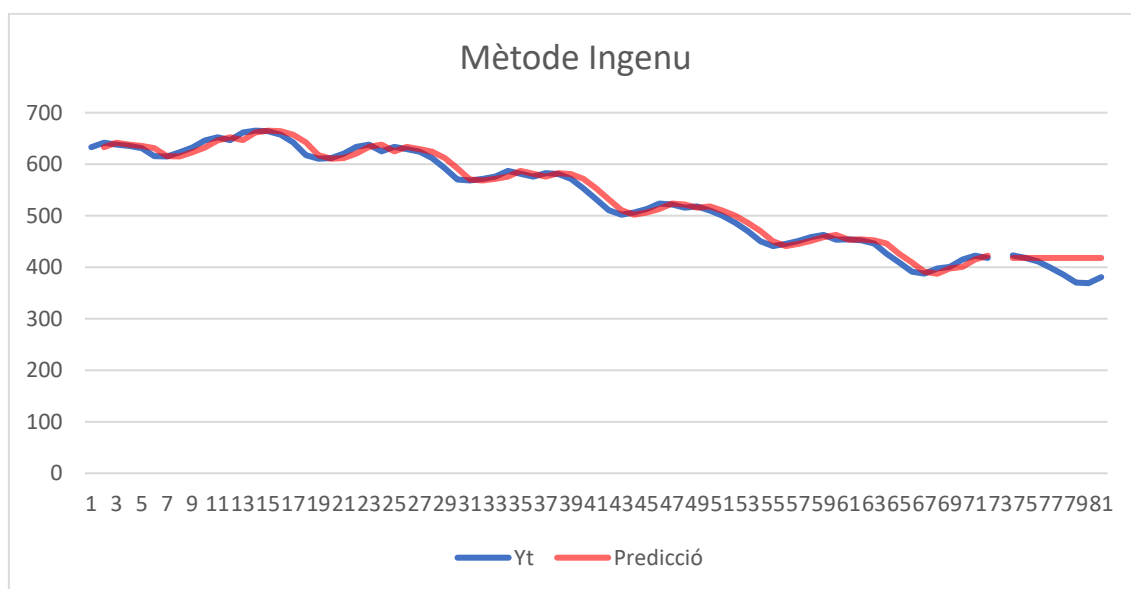
### 2.3.1. Mètode ingenu.

És el mètode més simple d'entre tots els mètodes de prediccions i consisteix en assignar al període  $t+1$  el valor observat en el període  $t$  en el període mostral i, en l'extra mostral, assignar sempre el valor del període  $T$  (mirar pestanya **1. Mètode Ingenu** a l'Excel).

Els resultats obtinguts són els següents:

Període Mostral			
EAM	EQM	EPAM	
8,904	115,703	1,69%	

Període Extra mostral			
EAM	EQM	EPAM	
24,638	943,526	6,49%	



Observem que l'EPAM del període extra mostral és major al 5%, la qual cosa significa que les prediccions que farem amb aquest mètode seran molt dolentes. Amb la representació gràfica o veiem més clarament: la línia vermella després del salt en blanc (la que correspon al període extra mostra) no s'ajusta bé a les dades observades.

Aquest mètode ens serà útil més endavant per a comparar els propers models de predicció ja que seran més sofisticats i, si ningun ens dona unes prediccions menys dolentes que les d'aquest mètode, podrem concloure que cap d'ells és adequat i s'hauria d'utilitzar un altre mètode complex.

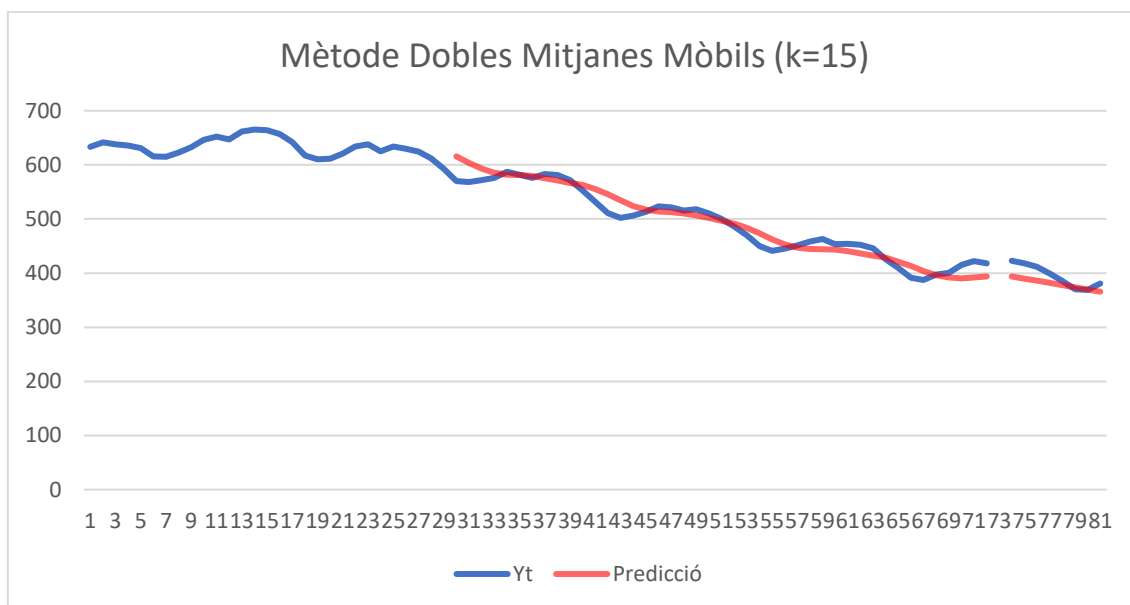
### 2.3.2. Mètode de les dobles mitjanes mòbils.

Amb aquest mètode s'obtenen les prediccions suposant que la tendència és localment lineal, lo que significa que la pendent es va definint a mesura que incorporem nova informació.

Els resultats obtinguts són (mirar càlculs a la pestanya 2. *Dobles mitjanes mòbils* a l'Excel):

Període Mostral			
EAM	EQM	EPAM	
14,293	311,541	2,97%	

Període Extra mostral			
EAM	EQM	EPAM	
15,842	359,302	3,90%	



S'ha fixat un alfa = 15 per tal que l'EPAM sigui inferior al 5%, això significa que estem utilitzant molta memòria (informació de temps anteriors). Ara veiem la línia de la predicció una mica millor que abans pel que fa al període extra mostral i que l'EPAM obtingut és 3.9%.

Les prediccions són menys dolentes que les d'abans ja que amb el mètode ingenu no utilitzàvem quasi memòria (com si fos k=1).

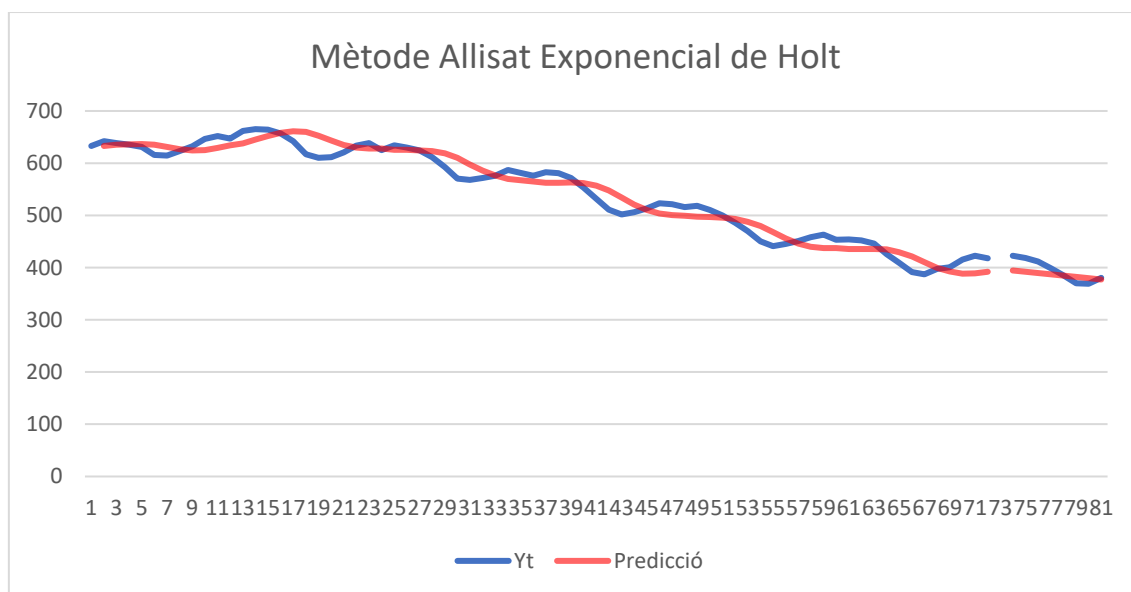
### 2.3.3. Mètode allisat exponencial de Holt (AEH).

En aquest mètode, igual que en l'anterior, es suposa que la tendència és localment lineal. Per tant, l'estimació de la tendència s'actualitza a mesura que s'incorpora informació però a diferència d'abans, la redefinició de la pendent es basa en tots els valors previs i no només en els k-valors previs.

Per a iniciar el mètode AEH cal suposar que el primer valor de tendència és el primer valor observat ( $\hat{t}_1 = y_1$ ) i que la pendent és 0 ( $\hat{\beta}_1 = 0$ ) ja que no tenim valors anteriors al primer.

Els resultats obtinguts són (mirar pestanya **3. Allisat exponencial de Holt** a l'Excel):

Període Mostral	EAM	EQM	EPAM
	16,448	381,353	3,15%
Període Extra mostral	EAM	EQM	EPAM
	14,319	293,344	3,54%



Per a obtenir aquest EPAM inferior al 5% (3.54%) s'ha hagut de fixar alfa i beta a 0.2, això significa que estem utilitzant molta memòria. Per definir aquets valors s'ha hagut de fer un joc de proves (anar canviant valors per alfa i beta entre 0 i 1, on 0 implica màxima memòria i 1, memòria mínima) i al final, la millor opció ha sigut aquesta.

Un altre cop, les prediccions amb aquest mètode sofisticat són millors que amb les del mètode ingenu, i ho comprovem tant amb els càlculs dels errors (tots són inferiors) com en la gràfica.



### 2.3.4. Mètode de la tendència lineal.

Amb el mètode de la tendència lineal, a diferència dels dos anteriors, es considera que la tendència és constant en tot el període mostral. S'estimen les prediccions aproximant la tendència mitjançant una funció lineal, la recta de regressió de tots valors del període mostral, és a dir que amb aquest mètode s'utilitza total la informació disponible.

Un cop obtinguts els coeficients de  $\beta_0$  i  $\beta_1$  mitjançant la regressió lineal (veure pestanya **Regressió** a l'Excel), les prediccions s'obtenen com:

$$\hat{y}_t = \hat{\beta}_0 + \hat{\beta}_1 * (t + 1) \text{ per al període mostral, i}$$

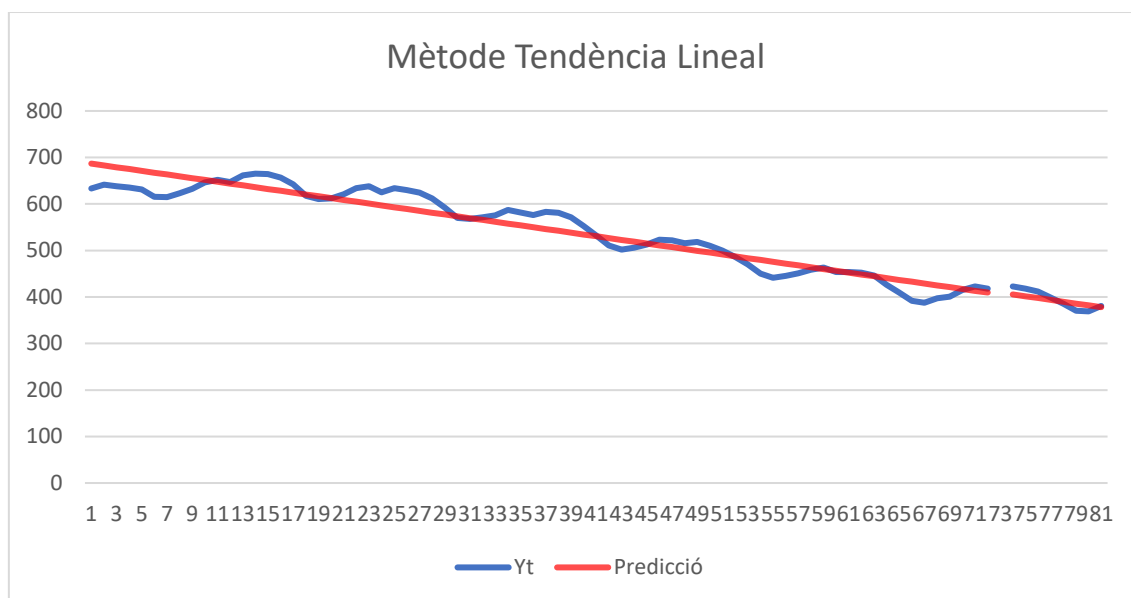
$$\hat{y}_t = \hat{\beta}_0 + \hat{\beta}_1 * (T + (t - T)) \text{ per al període extra mostral.}$$

Els resultats obtinguts són (veure pestanya **4. Tendència lineal** a l'Excel):

Període Mostral		
EAM	EQM	EPAM
20,881	653,370	3,79%

Període Extra mostral		
EAM	EQM	EPAM
11,081	155,026	2,79%



Al tenir les dades una tendència molt clara, la recta de prediccions per al període extra mostral s'ajusta força bé i si mirem l'EPAM, observem que és bo per a fer prediccions. En canvi, pel que fa a les prediccions del període mostral, les estimacions tenen un EQM i un EPAM més alt d'entre els 4 mètodes emprats.

### 3. Conclusions.

L'anàlisi realitzat ens permet concloure que la sèrie temporal sobre l'atur a Catalunya en els darrers 6 anys és una sèrie amb tendència però sense estacionalitat, és a dir, que les dades es distribueixen aleatòriament al voltant d'una tendència de decreixement.

Els mètodes més sofisticats utilitzats per a fer prediccions han sigut basats en la tendència lineal determinista i, a més, s'ha utilitzat un mètode simple (l'ingenu) per comprovar que els resultats obtinguts amb aquells mètodes eren els adequats.

D'entre tots, els mètode que ens ha ofert les estimacions menys dolentes és el de la tendència lineal a causa de la tendència tan marcada que tenen les nostres dades. Així, al tenir un error percentual absolut mig inferior al 3%, la capacitat predictiva és bona i al fer prediccions amb aquest mètode, aquestes seran correctes.