
LA REGRESSIONE LINEARE

— Per l'analisi di serie storiche —

Prof. Fabio Banderali (ITE Bassi - Lodi)



COSA SONO LE SERIE STORICHE

Le serie storiche sono **dati** che mostrano l'evoluzione di una grandezza rispetto al tempo.

Il tempo è considerato sempre come **variabile indipendente** (X), mentre la grandezza che si osserva come **variabile dipendente** (Y).

Lo scopo dell'analisi delle serie storiche è quello di prevedere l'andamento della grandezza rispetto al tempo, ovvero prevedere quali valori assumerà tale grandezza in futuro.

COSA SONO LE SERIE STORICHE

	A	B	C	D	E	F	G	H	I	J	K	L
1	ANNO	MILIONI DI FATTURATO										
2	2000	400										
3	2001	420										
4	2002	460										
5	2003	480										
6	2004	470										
7	2005	500										
8	2006	540										
9	2007	590										
10	2008	630										
11	2009	600										
12	2010	640										
13	2011	670										
14	2012	700										
15	2013	720										
16	2014	690										
17	2015	740										
18	2016	760										
19	2017	800										
20	2018	850										
21												

INDIVIDUA:

il dataset

la variabile
indipendente

la variabile
dipendente

COSA SONO LE SERIE STORICHE

VARIABLE INDIPENDENTE →

ANNO	MILIONI DI FATTURATO
2000	400
2001	420
2002	460
2003	480
2004	470
2005	500
2006	540
2007	590
2008	630
2009	600
2010	640
2011	670
2012	700
2013	720
2014	690
2015	740
2016	760
2017	800
2018	850

← **VARIABLE DIPENDENTE**

← **DATASET**

RAPPRESENTARE LE SERIE STORICHE

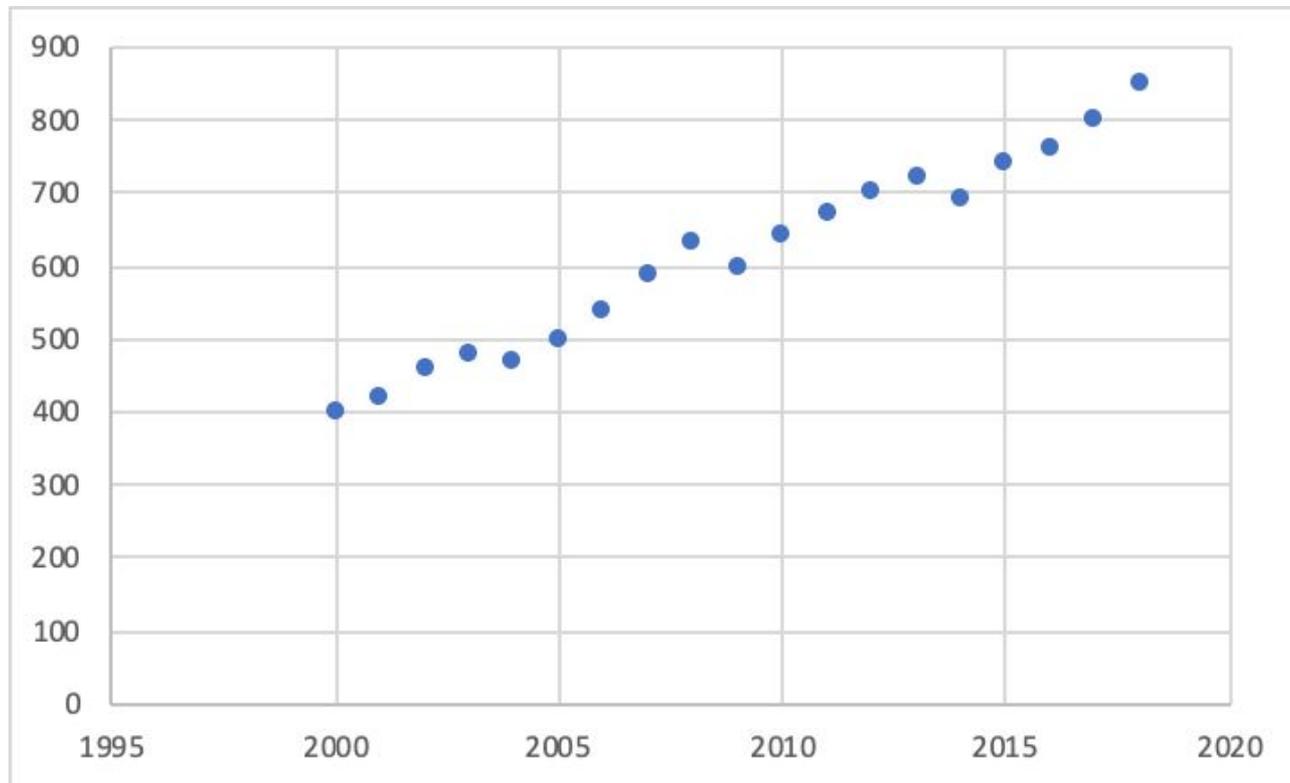
Per rappresentare graficamente l'andamento delle serie storiche si utilizza solitamente il **diagramma di dispersione**.

In **ascissa** viene riportato il tempo.

In **ordinata** viene riportata la grandezza osservata.

Il diagramma di dispersione permette a prima vista di capire se esiste una relazione fra le due variabili ovvero se è possibile osservare un **trend**.

RAPPRESENTARE LE SERIE STORICHE

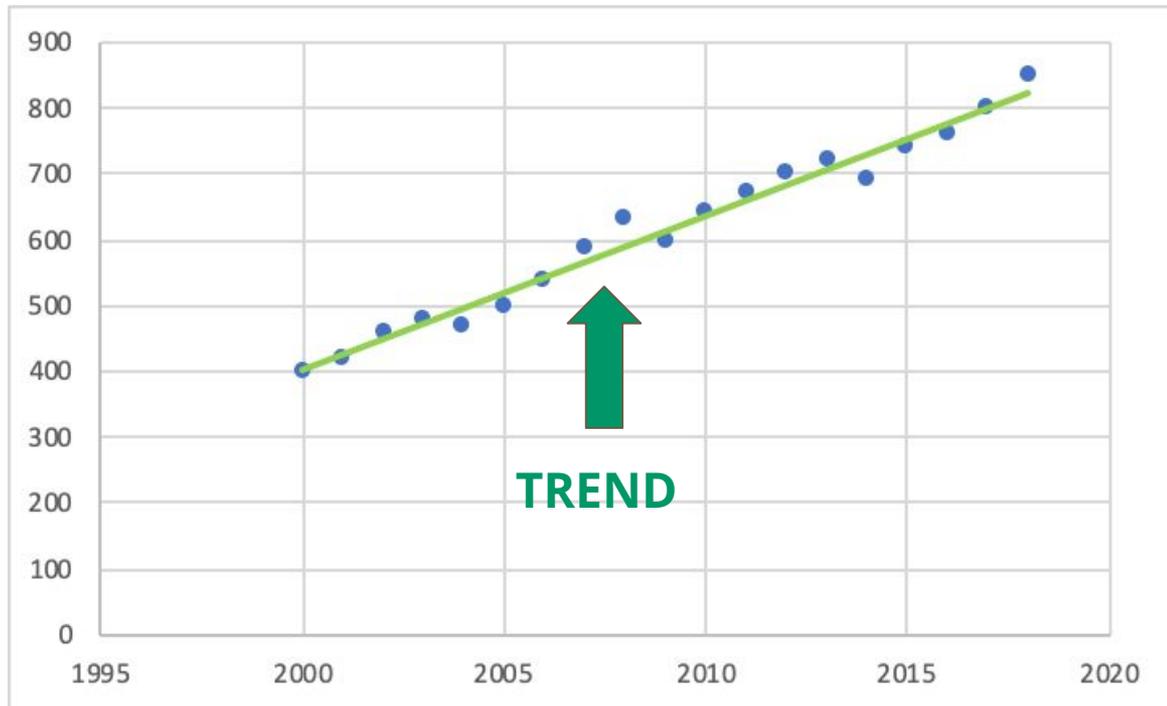


In quali anni il fatturato è aumentato?

In quali è diminuito?

Nel periodo osservato il fatturato è **generalmente** aumentato o diminuito? Overo quale **trend** ha avuto?

RAPPRESENTARE LE SERIE STORICHE



Il **trend** è rappresentato da una **retta** crescente ovvero il fatturato è generalmente aumentato con il passare del tempo.

L'equazione canonica della retta è:

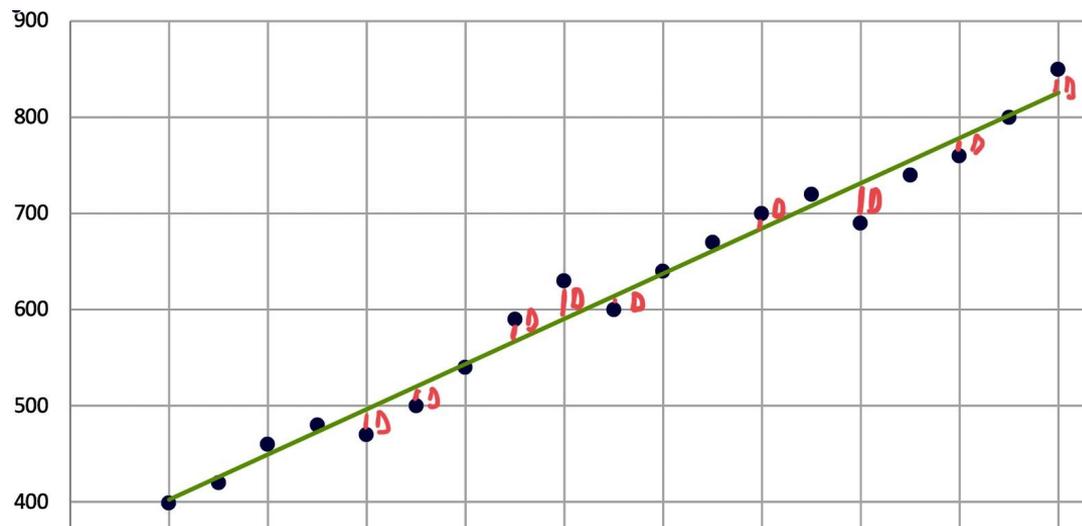
$Y = ax + b$ dove:
a: coefficiente angolare
b: intercetta

Come individuare il trend ovvero l'equazione della retta?

Il problema è trovare i valori di a e b che individuano la retta di trend.

LA RETTA DI REGRESSIONE

La retta del trend è detta anche **retta di regressione** ed è quella retta che rende minima la differenza (**D**) fra il valore osservato e l'ordinata della retta.



LA RETTA DI REGRESSIONE

La retta di regressione **approssima** dunque l'andamento della variabile rispetto al tempo. Come tutte le approssimazioni sono impliciti alcuni **errori di approssimazione** intesi come differenza fra il valore realmente assunto dalla grandezza e l'ordinata di un punto appartenente alla retta in corrispondenza di una certa ascissa.

D = errore di approssimazione della retta di regressione

LA RETTA DI REGRESSIONE

D = errore di approssimazione della retta di regressione

D = 0

l'ordinata del punto (anno) al quale ci si riferisce coincide con il valore reale della grandezza ovvero il punto giace sulla retta di regressione

D ≠ 0

è stato commesso un errore di approssimazione dalla retta di regressione. Il punto non giace sulla retta e la distanza fra il punto e la retta rappresenta appunto l'errore di approssimazione.

LA RETTA DI REGRESSIONE

La retta di regressione è quella retta che rende minimi la somma di tutte le differenze (D) elevate al quadrato; è necessario elevare al quadrato tutte le differenze per **evitare compensazioni di segno**.

Retta di regressione individua il coefficiente angolare **a** e l'intercetta **b** tali che

$$\min \sum D^2$$

Questa modalità di procedimento prende il nome di **metodo dei minimi quadrati**.

LA RETTA DI REGRESSIONE

La determinazione del coefficiente angolare e dell'intercetta della retta di regressione calcoli piuttosto laboriosi e alcune preconoscenze statistiche; pertanto utilizzeremo nel seguito la funzione **REGRESSIONE** di **EXCEL** imparando a utilizzare i risultati.



Qualora non presente lo strumento ANALISI DI DATI è necessario installarlo selezionandolo dai componenti aggiuntivi.

LA RETTA DI REGRESSIONE

OUTPUT OTTENUTO

	A	B	C	D	E	F	G	H	I	
1	OUTPUT RIEPILOGO									
2										
3	<i>Statistica della regressione</i>									
4	R multiplo	0,989094275								
5	R al quadrato	0,978307485								
6	R al quadrato corretto	0,977031455								
7	Errore standard	20,24004552								
8	Osservazioni	19								
9										
10	ANALISI VARIANZA									
11		<i>gdl</i>	<i>SS</i>	<i>MQ</i>	<i>F</i>	<i>Significatività F</i>				
12	Regressione	1	766,6804716	314077,8947	766,6804716	1,39067E-15				
13	Residuo	17	409,6594427	24,09761487						
14	Totale	18	1176,3400143							
15										
16		<i>Coefficienti</i>	<i>Errore standard</i>	<i>Stat t</i>	<i>Valore di significatività</i>	<i>Inferiore 95%</i>	<i>Superiore 95%</i>	<i>Inferiore 95,0%</i>	<i>Superiore 95,0%</i>	
17	Intercetta	-46544,94737	1703,160624	-27,32857178	1,72959E-15	-50138,30218	-42951,59255	-50138,30218	-42951,59255	
18	Variabile X 1	23,47368421	0,847762217	27,6889955	1,39067E-15	21,68506228	25,26230614	21,68506228	25,26230614	
19										

RETTA DI REGRESSIONE

$Y = 23,47 X - 46.544,95$

LA RETTA DI REGRESSIONE

OUTPUT OTTENUTO

	A	B	C	D	E	F	G	H	I
1	OUTPUT RIEPILOGO								
2									
3	<i>Statistica della regressione</i>								
4	R multiplo	0,989094275							
5	R al quadrato	0,978307485							
6	R al quadrato corretto	0,977031455							
7	Errore standard	20,24004552							
8	Osservazioni	19							
9									
10	ANALISI VARIANZA								
11		<i>gdl</i>	<i>SQ</i>	<i>MQ</i>	<i>F</i>	<i>Significatività F</i>			
12	Regressione	1	314077,8947	314077,8947	766,6804716	1,39067E-15			
13	Residuo	17	6964,210526	409,6594427					
14	Totale	18	321042,1053						
15									
16		<i>Coefficienti</i>	<i>Errore standard</i>	<i>Stat t</i>	<i>Valore di significatività</i>	<i>Inferiore 95%</i>	<i>Superiore 95%</i>	<i>Inferiore 95,0%</i>	<i>Superiore 95,0%</i>
17	Intercetta	-46544,94737	1703,160624	-27,32857178	1,72959E-15	-50138,30218	-42951,59255	-50138,30218	-42951,59255
18	Variabile X 1	23,47368421	0,847762217	27,6889955	1,39067E-15	21,68506228	25,26230614	21,68506228	25,26230614
19									

$R^2 =$ BONTA' APPROSSIMAZIONE

$0 \leq R^2 \leq 1$

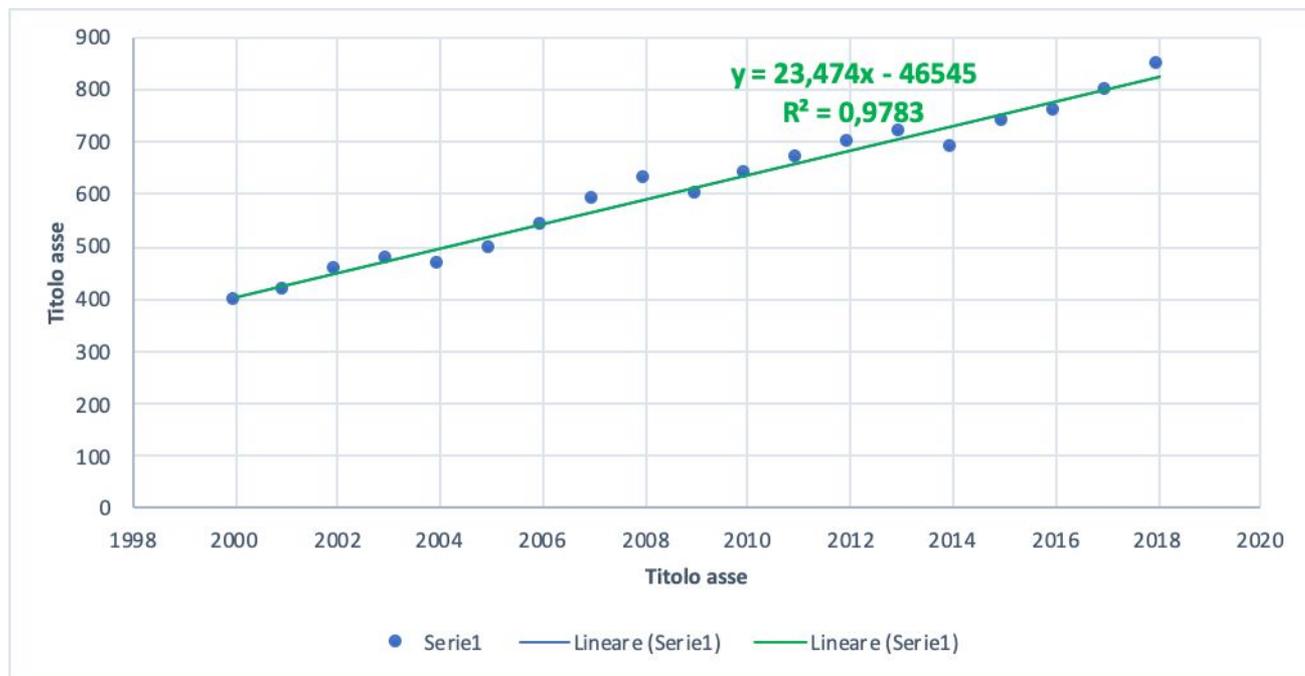
R=1 Ottimo

R=0 Pessimo

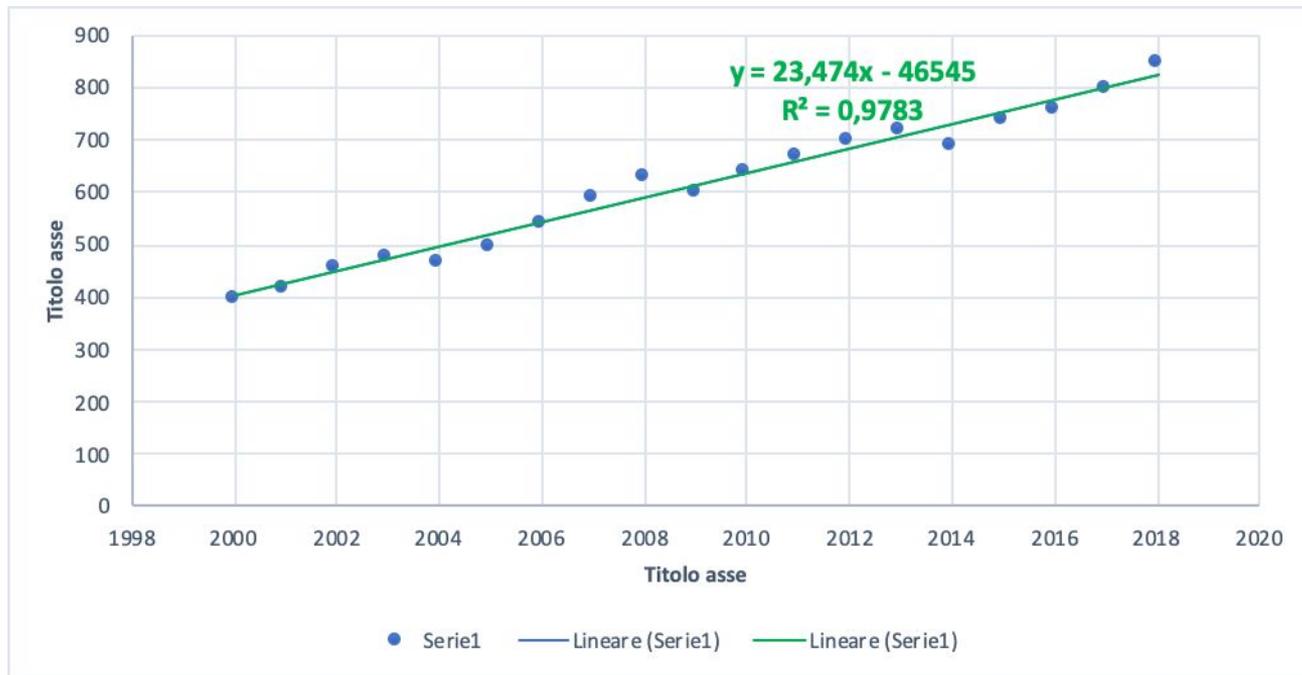


LA RETTA DI REGRESSIONE

Graficamente



LA RETTA DI REGRESSIONE



Utilizzando la retta di regressione costruita, effettua una previsione sul fatturato del 2019.

LA RETTA DI REGRESSIONE

Utilizzando la retta di regressione costruita, effettua una previsione sul fatturato del 2019.

Nell'equazione della retta di regressione

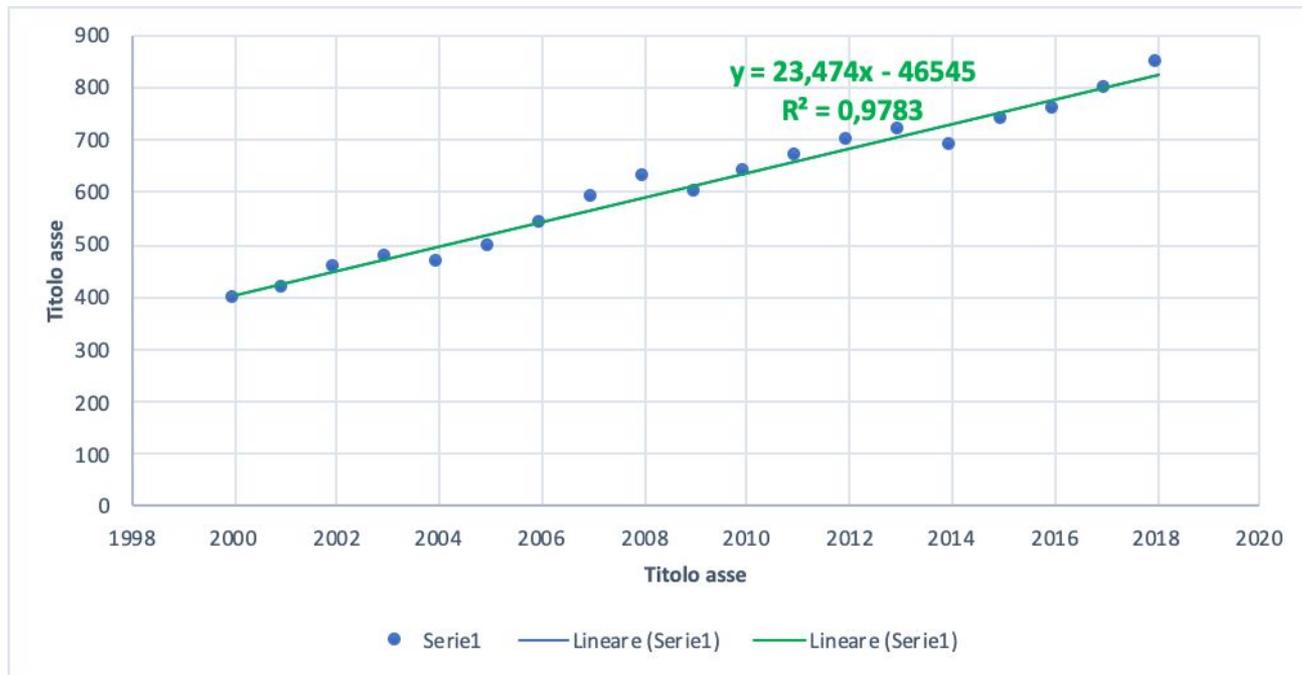
$$Y = 23,47 X - 46.544,95$$

sostituisco a X il valore 2019

$$Y (2019) = 23,47 * 2019 - 46.544,95 = 840,98$$

Per il 2019 si prevede un fatturato di 840,98 milioni di euro

LA RETTA DI REGRESSIONE



Calcolare il residuo (D) del 2008.

LA RETTA DI REGRESSIONE

Calcolare il residuo (D) del 2008.

Nel 2008 il fatturato conseguito è stato di 630 milioni di euro.

$$Y (2008) = 23,47 * 2008 - 46.544,95 = 582,81$$

La retta di regressione approssima il fatturato in 582,81 milioni di euro

L'errore di approssimazione, ovvero il residuo, è così determinato:

$$D (2008) = 630 - 582,81 = 47,19$$

PROVA TU

MESE	ORE DI STRAORDINARIO
1	28
2	33
3	37
4	40
5	36
6	38
7	42
8	47
9	50
10	53
11	57
12	55
13	58
14	61
15	64
16	69
17	73
18	71
19	74
20	77
21	80
22	84
23	90

Effettua una
previsione sulle
ore di
straordinario del
mese numero 30.

Calcola il residuo
del quinto mese