



Possibile ruolo della LEGGE DI BENFORD nella valutazione della qualità dei dati dei registri tumori

Emanuele Crocetti, Giorgia Randi, Tadek Dyba, Francesco Giusti, Carmen Martos, Roisin Rooney, Manola Bettio



ec.europa.eu/jrc



Joint Research Centre

the European Commission's
in-house science service

LEGGE DI BENFORD

123456789

2345

34567

45

567

0.6789

789000

891231

91

1456789

222345

0.0034

45

5

666789

79

89123

9686543726589487655987554323

0.000000000000000000000000021

PRIMA CIFRA SIGNIFICATIVA (FSD)

123456789

2345

34567

45

567

0.6789

789000

891231

91

1456789

222345

0.0034

45

5

666789

79

89123

9686543726589487655987554323

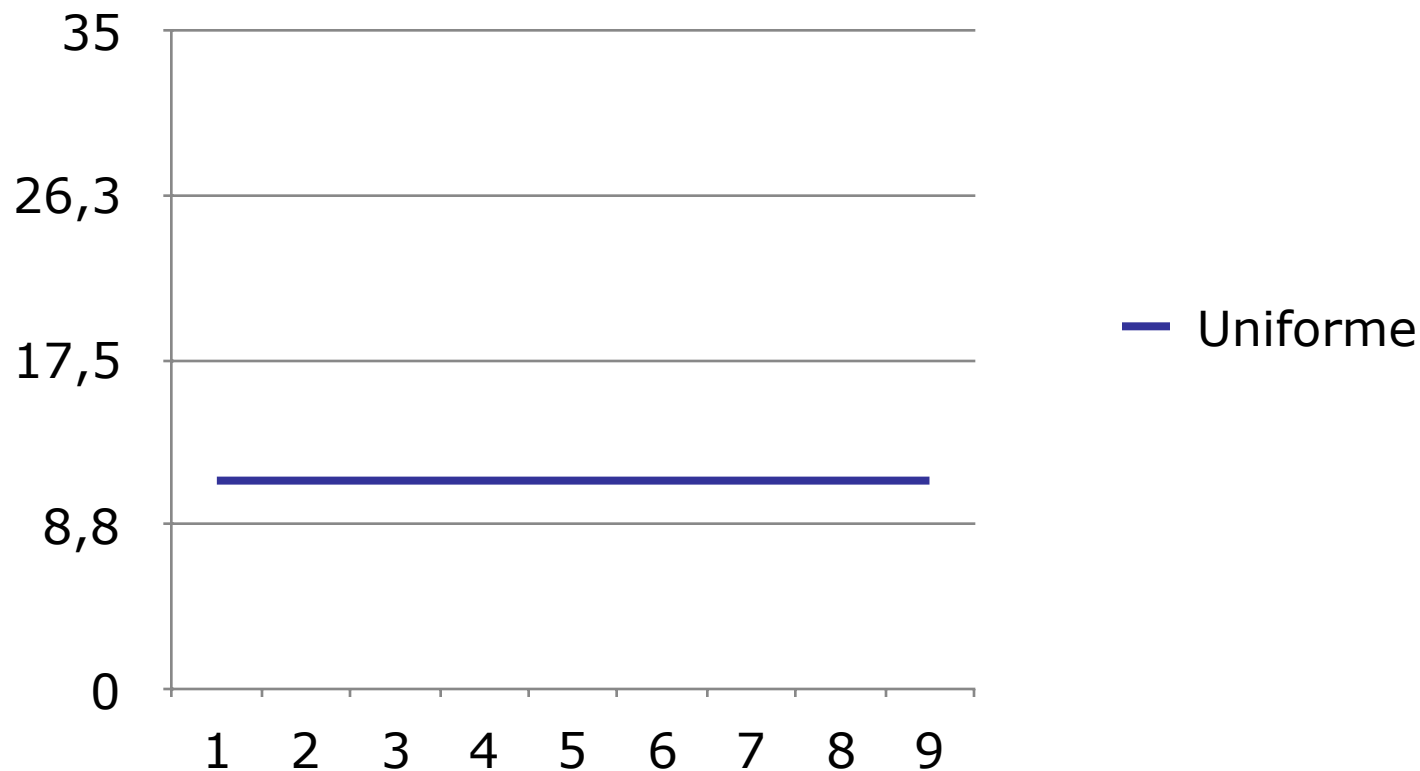
0.000000000000000000000000000021

LEGGE DI BENFORD

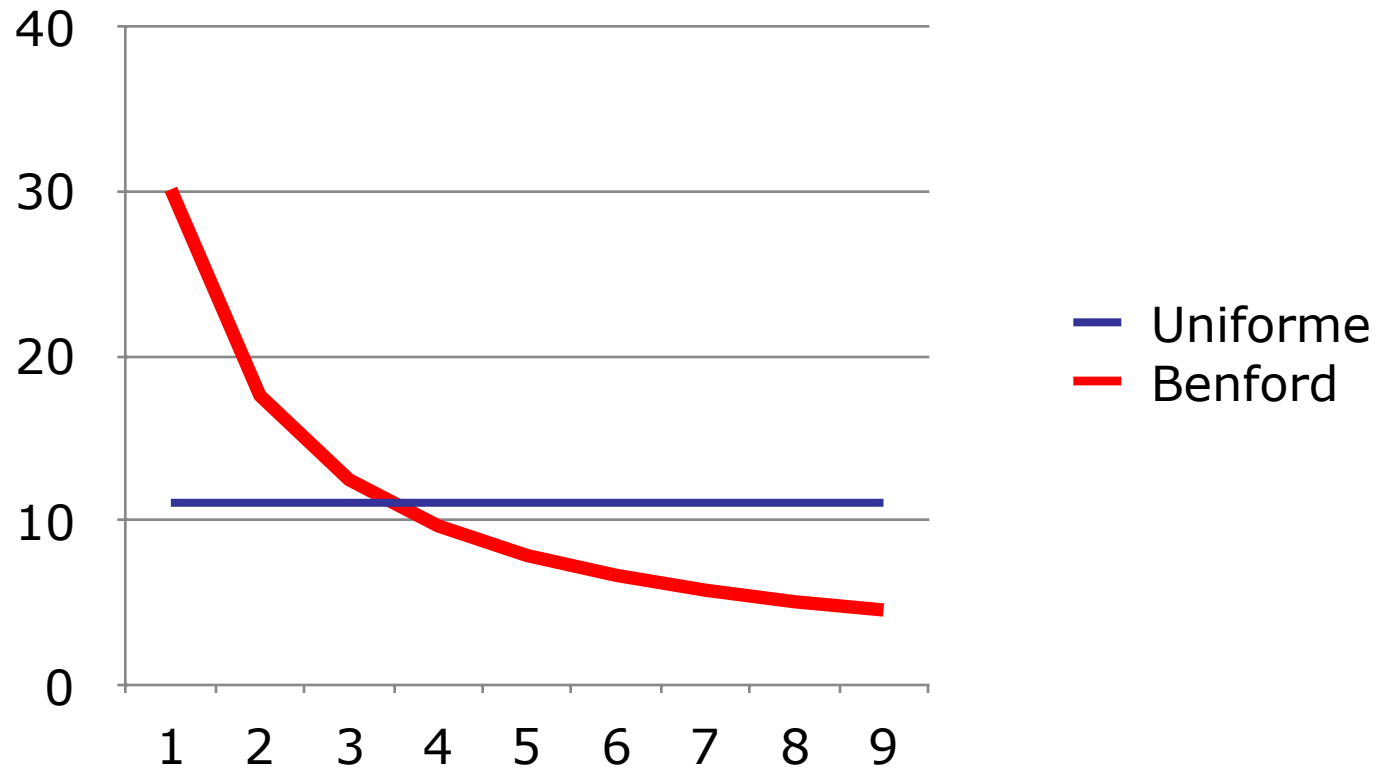
Descrive la probabilità che un numero (appartenente ad una serie numerica che rispetta la legge di Benford) inizi con:

1, 2, 3, 4, 5, 6, 7, 8, 9

Probabilità uniforme per un numero d'iniziare con 1,2,3,4,5,6,7,8,9



Distribuzione dei FSD secondo la legge di Benford e una probabilità uniforme



LEGGE DI BENFORD

Non tutti i numeri rispettano la legge di Benford, ma in quelli che lo fanno la distribuzione deve essere quella di Benford e le violazioni devono essere considerate con grande attenzione.

OBIETTIVO

I tassi d'incidenza seguono la legge di Benford?

Se la seguono, la distribuzione attesa della prima cifra dei tassi deve essere quella di Benford.

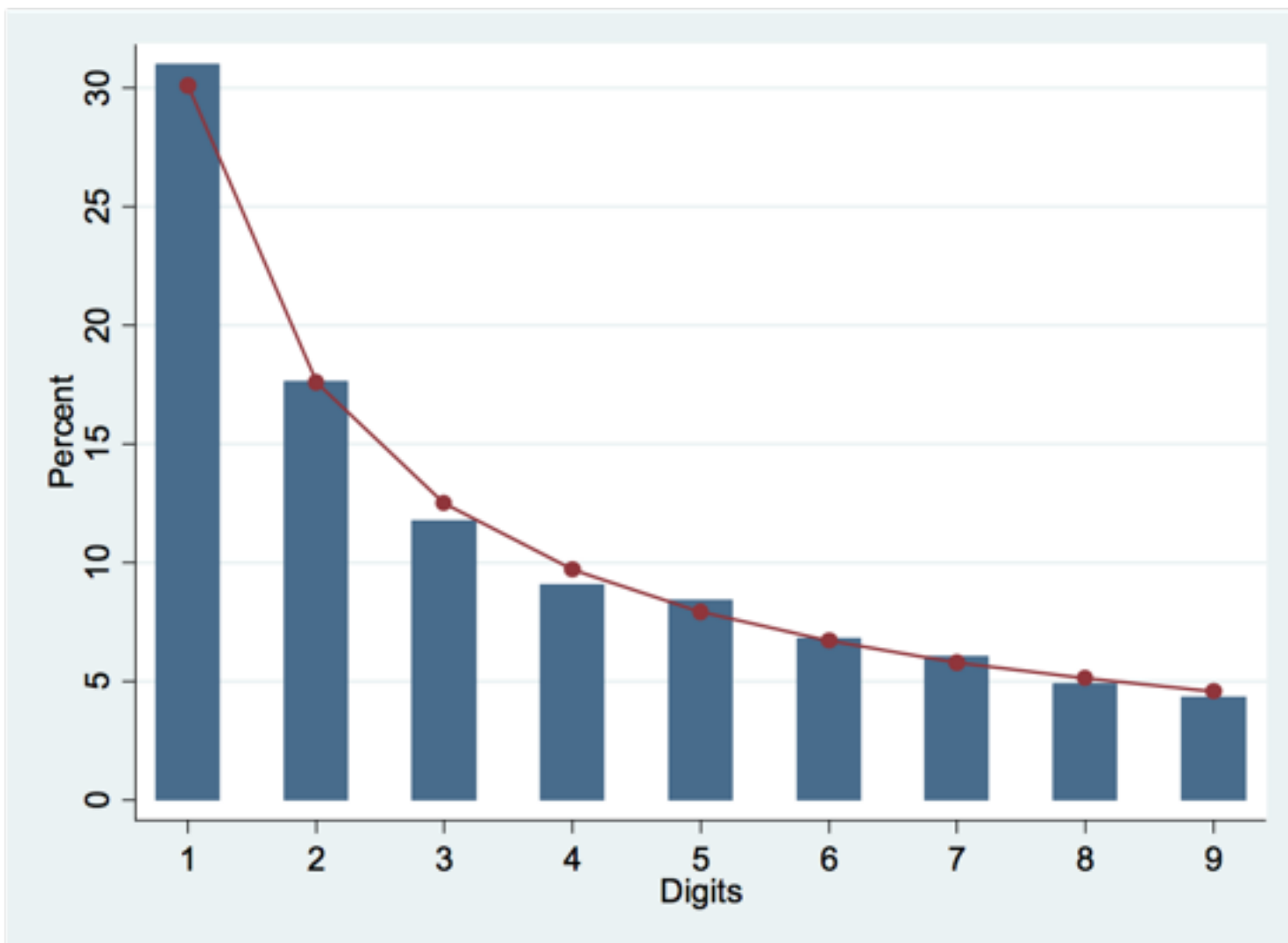
METODI

- da CI5C-X estratti i casi (dettagliati) e le popolazioni dei 33 Rt italiani
- calcolati i tassi grezzi per totale e RT.
- visualizzata la distribuzione del FSD
- applicate alcune statistiche per confrontare osservato e atteso

Perchè i tassi d'incidenza potrebbero rispettare la legge di Benford?

- Sono il prodotto di un calcolo fra due altri numeri
- Sono una serie ampia che occupa vari ordini di grandezza
- Non dipendono da scelte personali

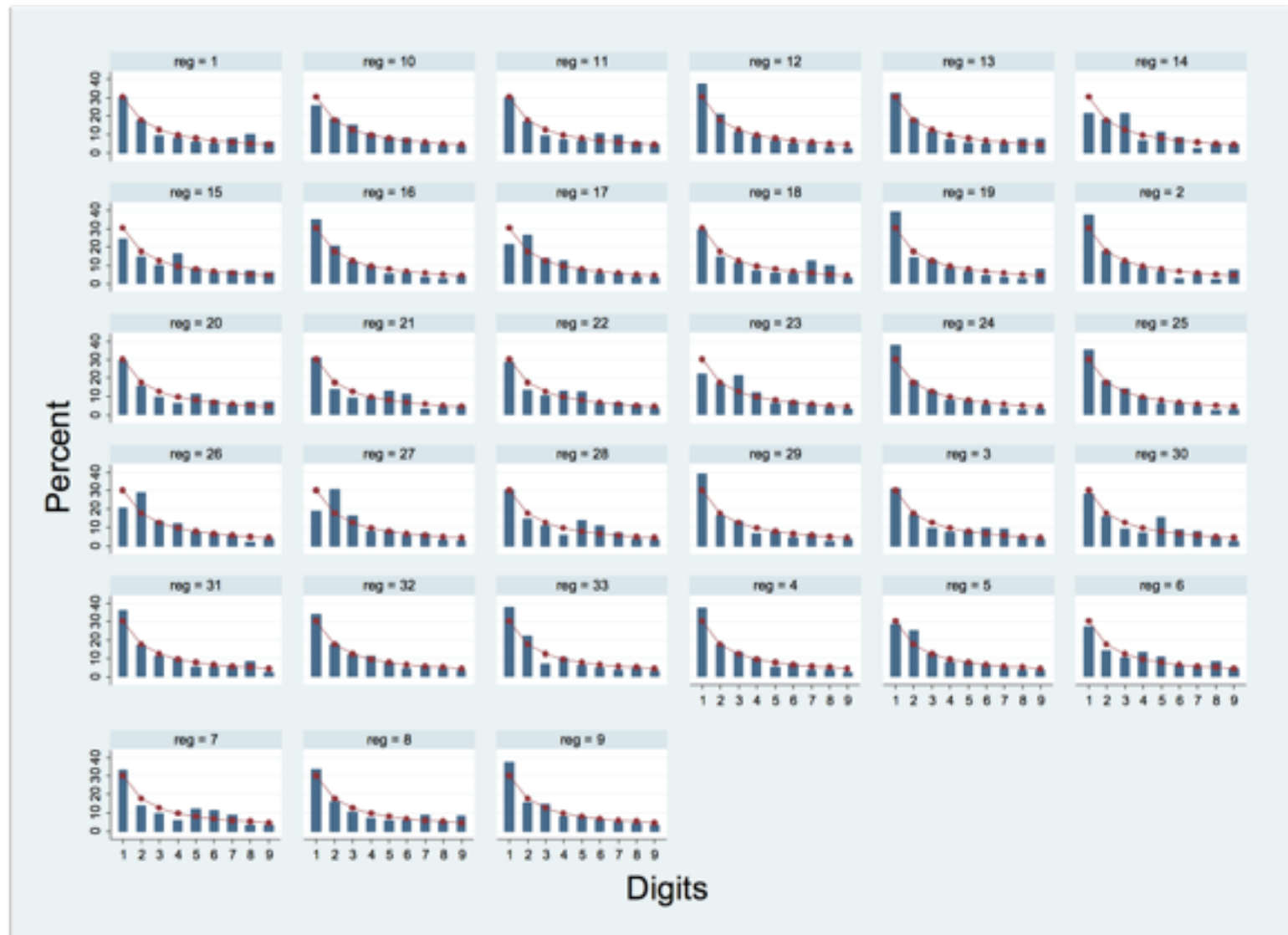
Sembra proprio che i tassi d'incidenza grezzi seguano la legge di Benford!



Cosa dice la statistica?

	n.	Chi2	VN*	r	m	d*	Chi2/ n	Z
Tutti i registri	97.724	165.1*	4.12*	0.999	0.009	0.014	0.002	1.22

La legge di Benford e i dati dei registri italiani



Indicatore di possibile non aderenza:

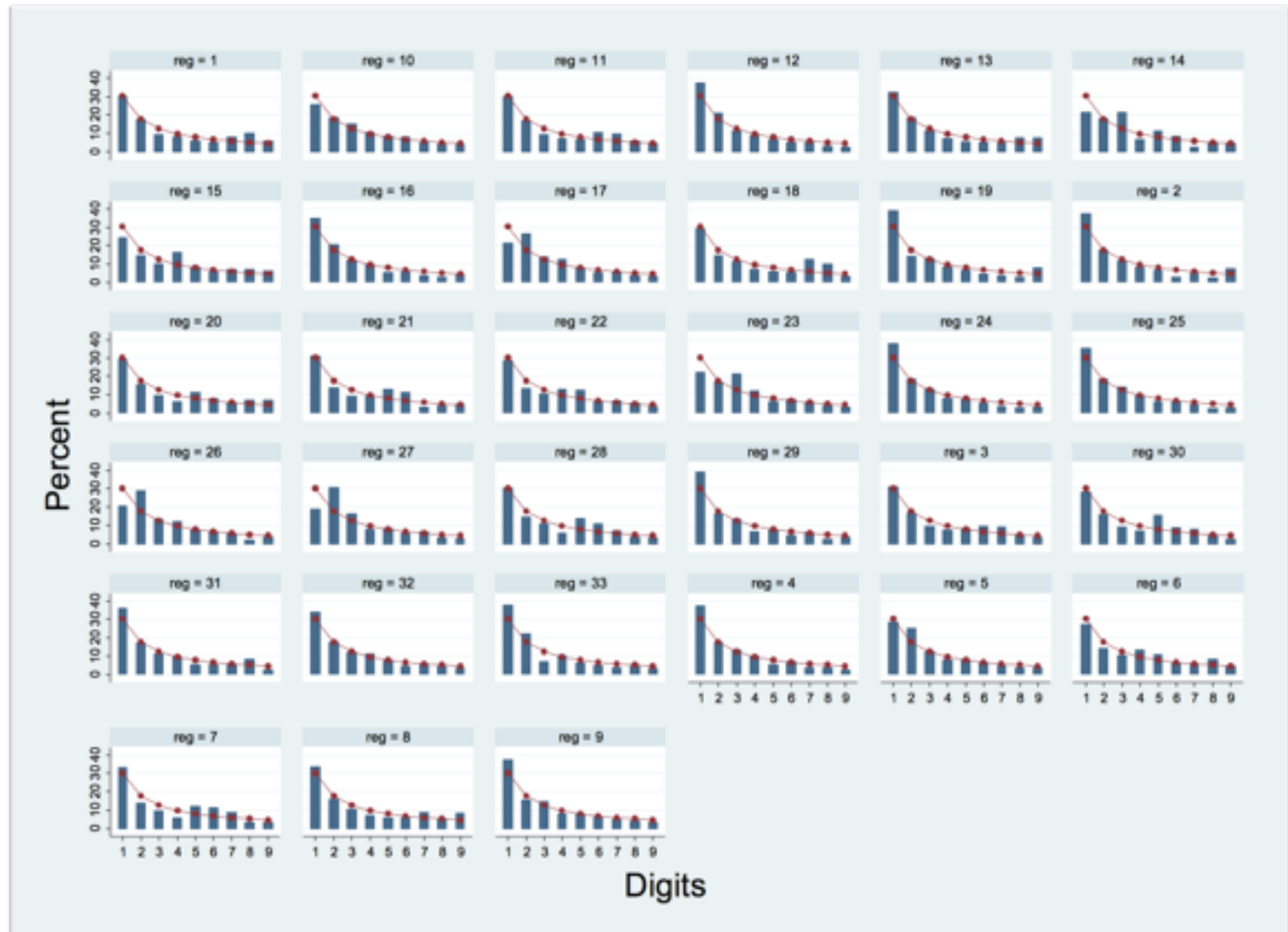
1 punto per zeta medio significativo

1 punto per r , d^* , m , e Chi^2/n se nel decile più estremo

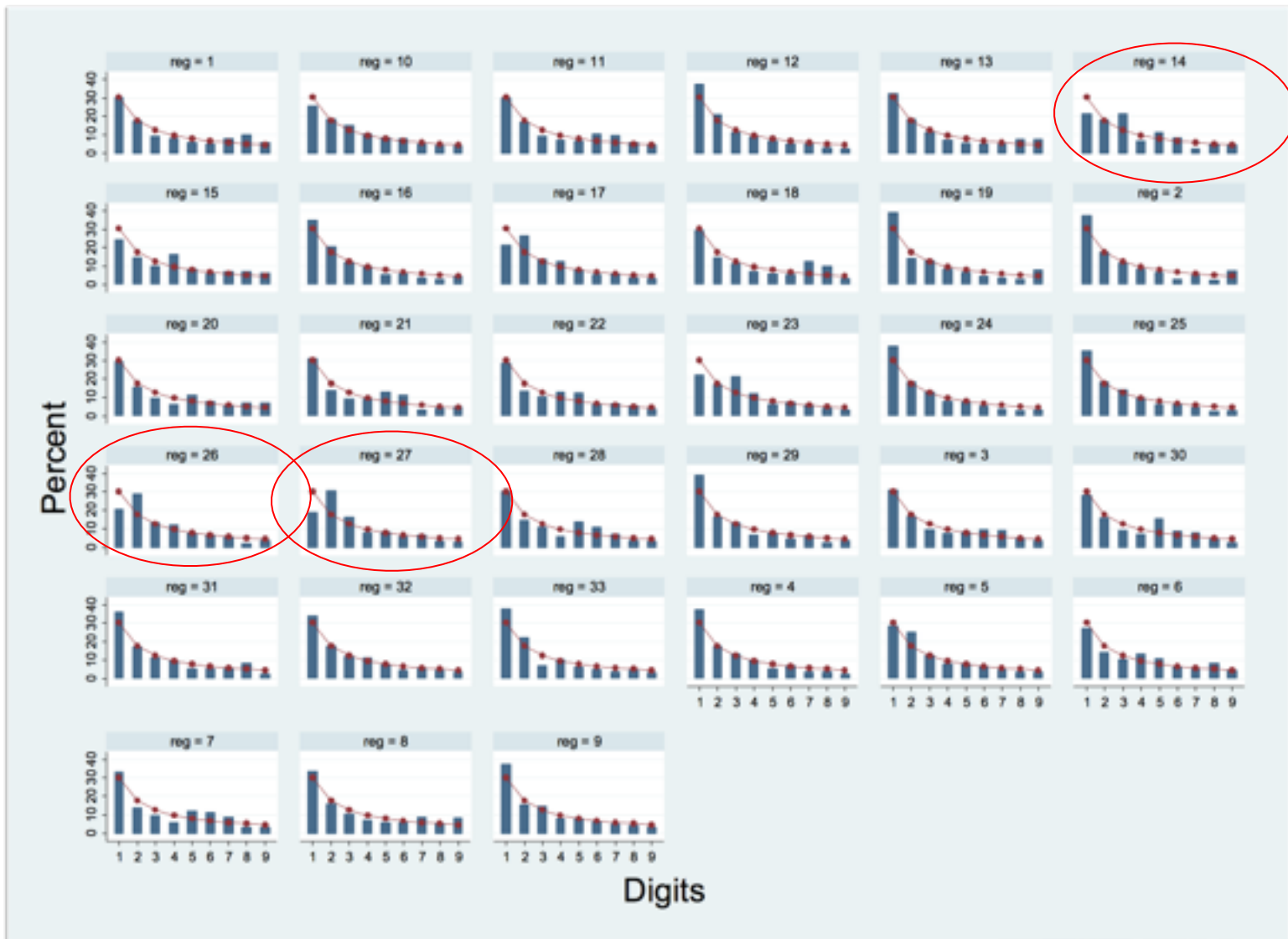
Indicatore di possibile non aderenza:

	r	m	d*	Chi2/ n	Z	indicatore
Registro Sud	1	1	1	1	1	5
Registro N-O	1	1	1	1	0	4
Registro Sud	1	1	1	0	0	3
Registro N-O	0	0	0	0	1	1
Registro N-O	0	0	0	1	0	1

La legge di Benford e i dati dei registri italiani



La legge di Benford e i dati dei registri italiani



CONCLUSIONI

I tassi d'incidenza seguono la legge di Benford.

Se un RT ha *sistematicamente* indicatori di distanza estremi (o statisticamente significativi) c'è un fondato sospetto di una violazione, che dovrebbe sottolineare la necessità di una valutazione accurata dei dati.

Verificare l'aderenza alla legge di Benford si propone come **un metodo nuovo, semplice e oggettivo** nel percorso di valutazione della qualità dei dati di un RT.

Stay in touch



JRC Science Hub:
ec.europa.eu/jrc



Twitter and Facebook:
@EU_ScienceHub



LinkedIn:
european-commission-joint-research-centre



YouTube:
JRC Audiovisuals



Vimeo:
Science@EC