

## Esercitazione 6 marzo 2014

### Esercizio 1 dal tema d'esame 10.02.2014 (parte prima)

Il gestore di un'azienda ha approvato il finanziamento di una medesima campagna pubblicitaria nei due stati europei in cui distribuisce i suoi prodotti.

Nella tabella successiva sono riportati i valori dell'ammontare delle vendite (in migliaia di euro) nello stato  $Y$  e nello stato  $Z$  in 10 settimane successive, ed il numero  $X$  di annunci pubblicitari trasmessi durante il fine settimana.

Settimana	$X$	$Y$	$Z$
1	2	50	46
2	5	57	53
3	1	41	48
4	3	54	40
5	4	54	42
6	1	38	42
7	5	63	60
8	3	48	51
9	4	59	52
10	2	46	43
<b>Totale</b>	<b>30</b>	<b>510</b>	<b>477</b>

1. Si calcolino media, mediana e varianza delle vendite nei due stati.
2. Si confronti la variabilità delle vendite in ciascuno stato con la variabilità del numero di annunci pubblicitari.
3. Si calcoli la correlazione tra le variabili  $X$  ed  $Y$  così come tra  $X$  e  $Z$ . In quale stato la campagna ha dato maggiori risultati? (*questa parte dell'esercizio sarà svolta in una lezione successiva*).

**Note.** Oltre alla mediana abbiamo calcolato anche il primo ed il terzo quartile. Ricordo le formule: indichiamo con  $n$  il numero di dati (qui  $n = 10$ ) e con  $p$  l'ordine del quartile (0.25, 0.50 e 0.75 per primo, secondo e terzo quartile rispettivamente).

- Se  $np$  è intero (p.es,  $np = 5$ ), allora si prende il punto medio tra il dato nella posizione  $np$  ed il dato nella posizione  $np+1$  del riordinamento crescente dei dati (nell'esempio, il punto medio tra il quinto ed il sesto dato del riordinamento crescente).
- Se  $np$  non è intero (p. es.  $np = 2.5$ ), si prende il dato nella prima posizione che segue al valore  $np$  nel riordinamento crescente dei dati (nell'esempio, il terzo dato del riordinamento crescente).

Abbiamo anche detto che la regola vale per un qualunque  $p$  tra 0 ed 1 (p. esempio, 10% o 95%).

### Esercizio 1 dal tema d'esame 4.09.2013 (parte prima)

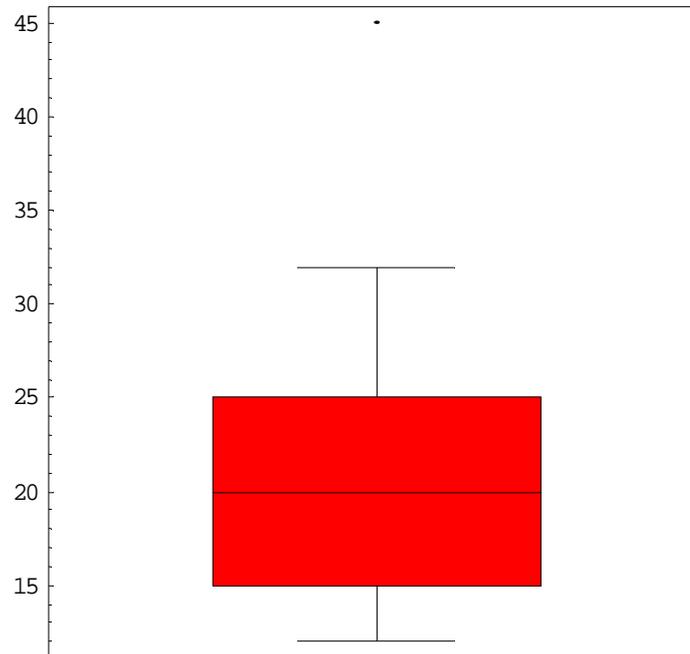
Si consideri la seguente distribuzione di frequenze:

$x_i$	12	15	16	20	21	24	25	30	32	45
$f_i$	8	12	3	5	2	5	5	1	8	1

1. Si calcolino la media e la mediana della distribuzione.
2. Si discuta la simmetria della distribuzione. (*solo accennato, ci torniamo in una delle prossime lezioni*)

3. Si rappresenti la variabilità della distribuzione tramite un grafico opportuno.

**Soluzione** del punto 3. è il *box-plot* qui sotto.



Si è discusso della simmetria dei dati a partire dal *box-plot* e dell'importanza di mettere in luce i valori anomali (o *outlier*).

### **Esercizio**

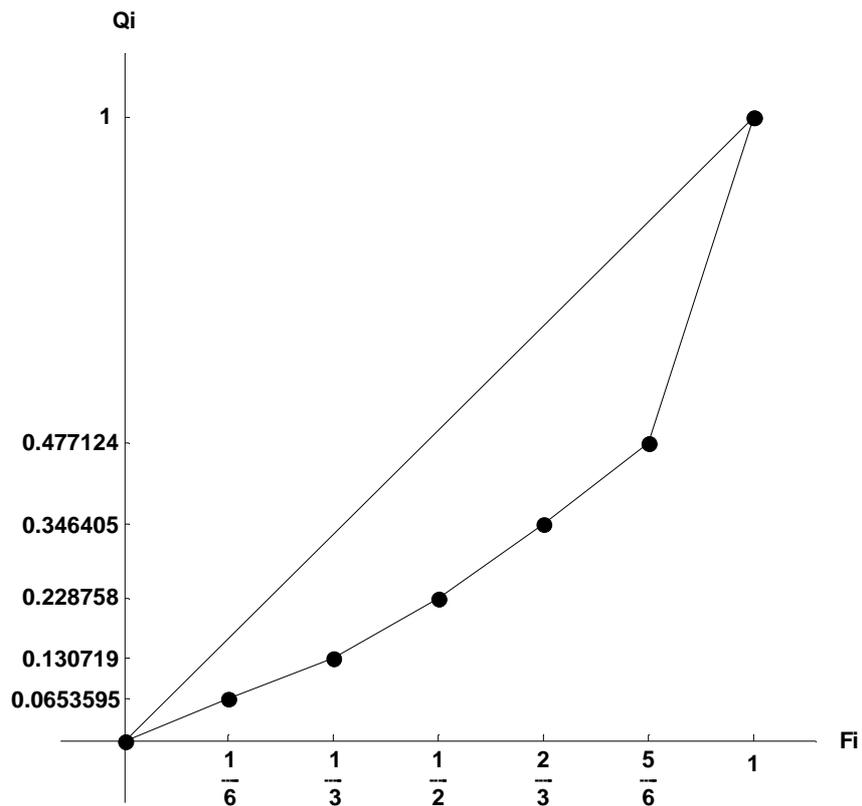
La tabella 1 riporta il numero (in milioni) di telespettatori sintonizzati sulle tre reti ammiraglie della RAI e sulle tre reti ammiraglie di Mediaset durante la prima serata del 21.02.2014.

**Tabella 1 – Dati originali**

<b>Rete</b>	<b>N. di telespettatori (in milioni)</b>	<b>Programma</b>
RAI 1	8	Sanremo
RAI 2	1	Virus
RAI 3	1.5	Film TV
Canale 5	1	Matrix
Italia 1	2	Arrow
Rete 4	1.8	Quarto grado
<b>Totale</b>	<b>15.3</b>	

Si rappresenti tramite una curva di Lorenz la concentrazione dei telespettatori nelle 6 reti e si calcoli l'indice di concentrazione di Gini.

**Soluzione.** La curva è rappresentata nella figura seguente.



**Figura 1 – Curva di concentrazione per i dati in Tabella 1**

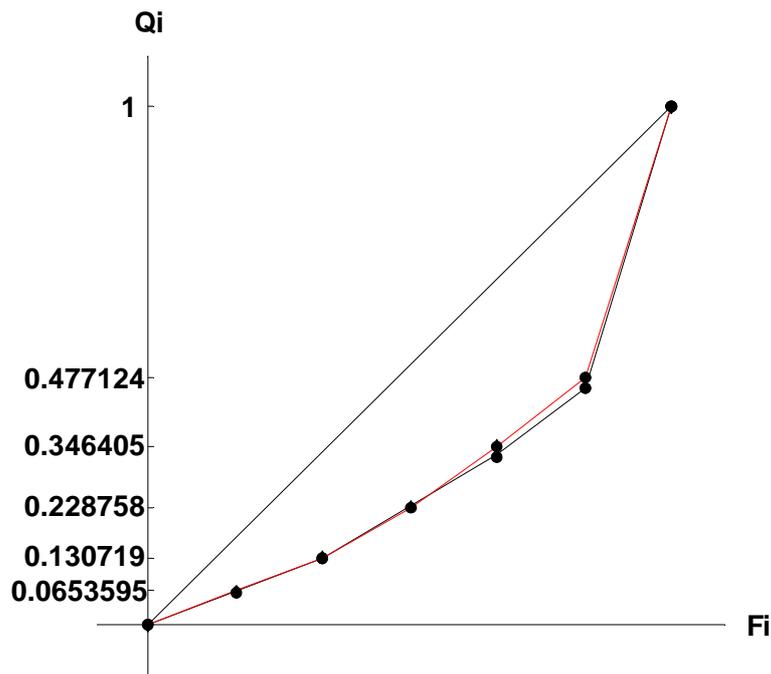
Indice di Gini: 0.494, corrispondente ad una concentrazione media (si ricordi che l'indice varia da 0, minima concentrazione, a 1, massima concentrazione).

**Note.** Si è discusso come si possa determinare allo stesso modo la curva di Lorenz e l'indice di Gini anche per dati che si presentino in una tabella di frequenza. Ce ne siamo inventata una che non ricordo esattamente, ma che potrebbe essere questa di Tabella 2:

**Tabella 2 – Tabella di frequenze discussa in classe (o simile...)**

Frequenza	N. di telespettatori (in milioni)
2	1
2	1.5
1	2
1	8
<b>Totale</b>	<b>15</b>

Rappresento nella figura che segue la curva di concentrazione ottenuta da questa tabella di frequenza (curva rossa) insieme a quella precedentemente ottenuta. Non c'è molta differenza ma ce l'aspettavamo perché i dati sono sostanzialmente simili.



**Figura 2 – Curva di Lorenz per i dati in Tabella 1 (nera) a confronto con la curva per i dati in Tabella 2 (rosso).**

Aggiungo, infine, la curva di concentrazione per un'altra tabella di frequenza, pure accennata in classe:

**Tabella 3 – Variazione della tabella 2.**

<b>Frequenza</b>	<b>N. di telespettatori (in milioni)</b>
2	1
2	1.5
1	5
1	5
<b>Totale</b>	<b>15</b>

Nella figura 3 confrontiamo la curva di Lorenz per i dati in Tabella 2 (rossa) con quella per i dati in Tabella 3 (blu). La curva blu si avvicina maggiormente alla bisettrice del primo quadrante, come ci si aspetta perché i telespettatori delle due reti più seguite in Tabella 2 sono stati ugualmente suddivisi passando alla Tabella 3.

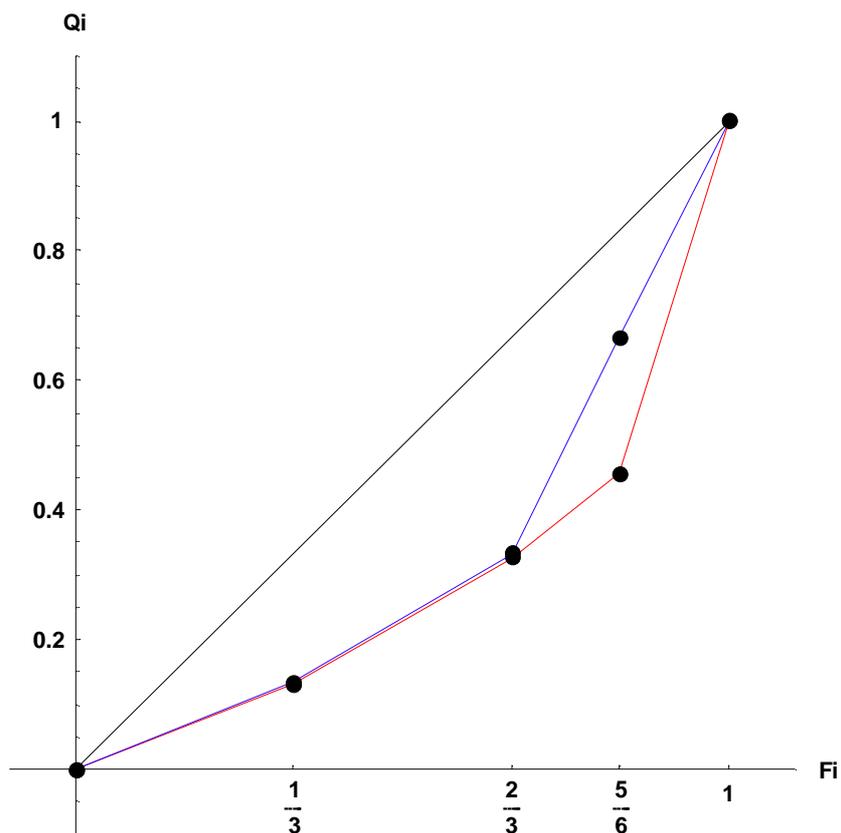


Figura 3 - Curva di Lorenz per i dati in Tabella 2 (rosso) a confronto con la curva per i dati in Tabella 3 (blu).

**Esercizio**

La tabella riporta la suddivisione dei dipendenti per qualifica in due aziende dello stesso settore. Rappresentare i dati con un grafico opportuno e confrontare l'eterogeneità delle due suddivisioni.

Qualifica	Azienda A	Azienda B
Operaio	10	90
Impiegato	6	45
Dirigente	2	15
<b>Totale</b>	<b>18</b>	<b>150</b>