Statistica
Esercitazione IV

13 Ottobre 2022

Esercizio 1 (Indice di eterogeneità del Gini, varianza, coefficiente di variazione) La seguente tabella a doppia entrata riporta il "liceo di provenienza" (X) e il "tempo in minuti dedicato alla visita di una mostra" (Y) di 80 studenti intervistati.

$X \backslash Y$	$5 \dashv 10$	$10 \dashv 20$	$20 \dashv 25$
Classico	12	12	6
Scientifico	38	2	10

1. Considerate le distribuzioni condizionate di $X|y \in (5;10]$ e $X|y \in (20;25]$ si proceda a:
   (a) rappresentarle graficamente;
   (b) determinare gli opportuni indici di posizione;
   (c) calcolare gli opportuni indici di variabilità (Indice di eterogeneità del Gini) e confrontarli.
2. Considerate le distribuzioni condizionate di $Y|x = \textit{Classico}$ e $Y|x = \textit{Scientifico}$ si proceda a:
   (a) rappresentarle graficamente;
   (b) determinare gli opportuni indici di posizione;
   (c) calcolare gli opportuni indici di variabilità (varianza e coefficiente di variazione) e confrontarli.

Esercizio 2 (Varianza e coefficiente di variazione di trasformazioni lineari di una variabile statistica) Data una variabile statistica X di media $\mu = 2$ e varianza $\sigma^2 = 6.0516$, si consideri la variabile $Y = 4 + 7X$. Dunque, calcolare la media, varianza e coefficiente di variazione di Y.

Esercizio 3 (Teorema di scomposizione della varianza) Si consideri la tabella a doppia entrata dell'Esercizio 1, in cui sono riportati il "liceo di provenienza" (X) e il "tempo in minuti dedicato alla visita di una mostra" (Y) di 80 studenti intervistati.

1. Calcolare la varianza di Y.

2. Calcolare la varianza di Y usando il teorema di scomposizione della varianza, utilizzando le distribuzioni condizionate di Y già ottenute precedentemente.

Esercizio 4 (Box-Plot) Costruire il Box-Plot per la seguente serie statistica:

x_i	n_i
18	4
23	40
26	36
29	70
32	50
	200

Esercizio 5 (Simmetria e indice di Fisher) Si consideri la seguente tabella che riporta la distribuzione di frequenza della variabile X:

X	1	a	3
f_i	0.2	b	c

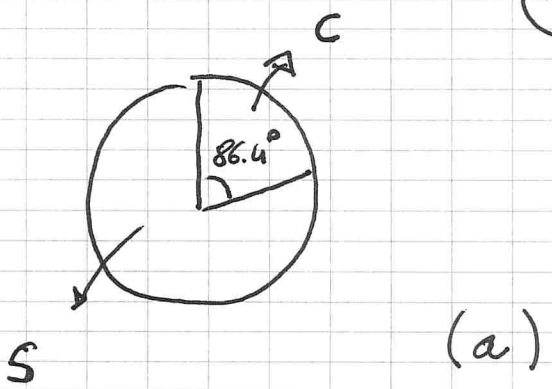
1. Si indichino i valori che devono assumere a, b e c affinché la distribuzione sia simmetrica.
2. Si calcolino media, varianza e indice di Fisher (detto anche di asimmetria o di skewness) per la variabile X, commentando i risultati.
3. Siano $Y = 4 + 2X$ e $Z = 3X$. Determinare media, varianza e indice di Fisher di Y e Z.

ESERCIZIO 1

1.

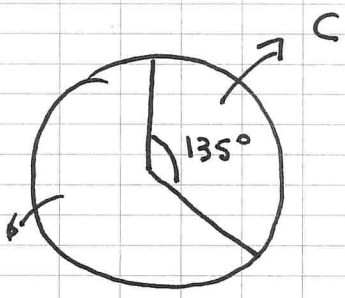
VARIABILE QUALITATIVA SCONNESSA

$x_i \mid y \in (5;10]$	n_i	f_i	$\alpha_i = 360° \cdot f_i$
CLASSICO	18	0.24	86.4°
SCIENTIFICO	38	0.76	273.6°
	50	1	360°



(a)

$x_i \mid y \in (20;25]$	n_i	f_i	α_i
CLASSICO	6	0.375	135
SCIENTIFICO	10	0.625	225
	16	1	360°



b) DATO QUALIT. SCONNESSO: SOLO "MODA"

$$\text{Moda}(X \mid y \in (5;10]) = \text{"SCIENTIFICO"} = \text{Moda}(X \mid y \in (20;25])$$

c) DATO QUALITATIVO: INDICE DI GINI

$$E(X \mid y \in (5;10]) = 1 - (0.24^2 + 0.76^2) = 0.3648$$
$$E(X \mid y \in (20;25]) = 1 - (0.375^2 + 0.625^2) = 0.4688$$
$$E_{MAX} = 1 - \frac{1}{k} = 1 - \frac{1}{2} = 0.5$$

$$E_N(X \mid y \in (5;10]) = 0.3648 / 0.5 = 0.7296$$
$$E_N(X \mid y \in (5;10]) = 0.4688 / 0.5 = 0.9376$$

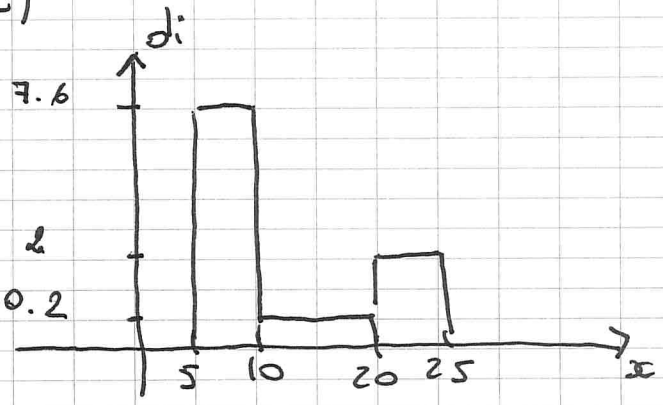
$E_N(X|y \in (20;25]) > E_N(X|y \in (5;10]) \Rightarrow X|y \in (20;25]$ HA MAGGIORE (e)
VARIABILITÀ DI $X|y \in (5;10]$. INOLTRE, $X|y \in (20;25]$ È CARATTERIZZATA DA
ELEVATA VARIABILITÀ, POICHÉ $E_N(X|y \in (20;25]) = 0.9376$ CHE È VICINO AD 1.

2.

$Y\|x$ = CLASSICO	y_i^c	n_i	a_i	$d_i = \frac{n_i}{a_i}$	f_i	F_i
5 – 10	7.5	12	5	2.4	0.4	0.4
10 – 20	15	12	10	1.2	0.4	0.8 ←
20 – 25	22.5	6	5	1.2	0.2	1
		30			1	

(a)

$y\|x$ = SCIENTIFICO	y_i^c	n_i	a_i	d_i	f_i	F_i
5 – 10	7.5	38	5	7.6	0.76	0.76 ←
10 – 20	15	2	10	0.2	0.04	0.8
20 – 25	22.5	10	5	2	0.2	1
		50				



b) $Y|X=$ CLASSICO

- CLASSE MODALE : $(5;10] \rightarrow$ Moda : 7.5
- CLASSE MEDIANA : $(10;20]$

Mediana : $h_{i-1} + (0.5 - F_{i-1}) \cdot \frac{a_i}{f_i}$

$= 10 + (0.5 - 0.4) \cdot \frac{10}{0.4} = 12.5$

- $M(Y|X=\text{CLAS.}) = 7.5 \cdot 0.4 + 15 \cdot 0.4 + 22.5 \cdot 0.2$
$= 13.5$

c) $M(Y^2|X=\text{CLASS.}) = 7.5^2 \cdot 0.4 + 15^2 \cdot 0.4 + 22.5^2 \cdot 0.2$
$= 213.75$

$\text{Var}(Y|X=\text{CLAS.}) = 213.75 - 13.5^2 = 31.5$

$CV(Y|X=\text{CLASS}) = \frac{\sqrt{31.5}}{13.5} = 0.4157$

$Y|X=$ SCIENTIFICO

- CLAS. MODALE: $(5;10] \rightarrow$ Moda : 7.5
- CLASSE MEDIANA : $(5;10]$

Mediana : $5 + (0.5 - 0) \cdot \frac{5}{0.76}$

$= 8.2895$

- $M(Y|X=\text{SC.}) = 7.5 \cdot 0.76 + 15 \cdot 0.04 + 22.5 \cdot 0.2$
$= 10.80$

$M(Y^2|X=\text{SC.}) = 7.5^2 \cdot 0.76 + 15^2 \cdot 0.04 + 22.5^2 \cdot 0.2$
$= 153$

$\text{Var}(Y|X=\text{SC.}) = 153 - 10.8^2 = 36.36$

$CV(Y|X=\text{SC.}) = \frac{\sqrt{36.36}}{10.80} = 0.5583$

$<$

$Y|X=$ SC. È CARATTERIZZATA DA MAGGIORE VARIABILITÀ.

ESERCIZIO 2

X: $\mu = 2$, $\sigma^2 = 6.0516$; $Y = 4 + 7X$

- $M(Y) = M(4 + 7X) = 4 + 7M(X) = 4 + 7 \cdot 2 = 18$
- $Var(Y) = Var(4 + 7X) = Var(7X) = 7^2 Var(X) = 49 \cdot 6.0516 = 296.5284$

$$Y = a + bX \Rightarrow CV(Y) = \frac{CV(X)}{\left| 1 + \frac{a}{bM(X)} \right|}$$

- $CV(X) = \frac{\sqrt{6.0516}}{2} = 1.23$

$$CV(Y) = \frac{1.23}{\left| 1 + \frac{4}{7 \cdot 2} \right|} = 0.9567$$

ESERCIZIO 3

1.

y_i	y_i^c	n_i	f_i
5–10	7.5	50	0.625
10–15	15	14	0.175
20–25	22.5	16	0.2
		80	1

$M(Y) = 7.5 \cdot 0.625 + 15 \cdot 0.175 + 22.5 \cdot 0.2 = 11.8125$

$M(Y^2) = 7.5^2 \cdot 0.625 + 15^2 \cdot 0.175 + 22.5^2 \cdot 0.2 = 175.7813$

$Var(Y) = 175.7813 - 11.8125^2 = 36.2461$

2. $n_c = 30$; $M(Y|x=c) = 13.5$, $Var(Y|x=c) = 31.5$

$n_s = 50$; $M(Y|x=s) = 10.8$, $Var(Y|x=s) = 36.36$

- $M(Y) \overset{\text{PROP. ASS.}}{=} \dfrac{13.5 \cdot 30 + 10.8 \cdot 50}{30+50} = 11.8125 =: \mu$

- $\sigma_B^2 = \sum_{j=1}^{k} (\mu_{Y|x_j} - \mu)^2 \cdot \dfrac{n_j}{n} = (13.5 - 11.8125)^2 \cdot \dfrac{30}{80} + (10.8 - 11.8125)^2 \cdot \dfrac{50}{80}$

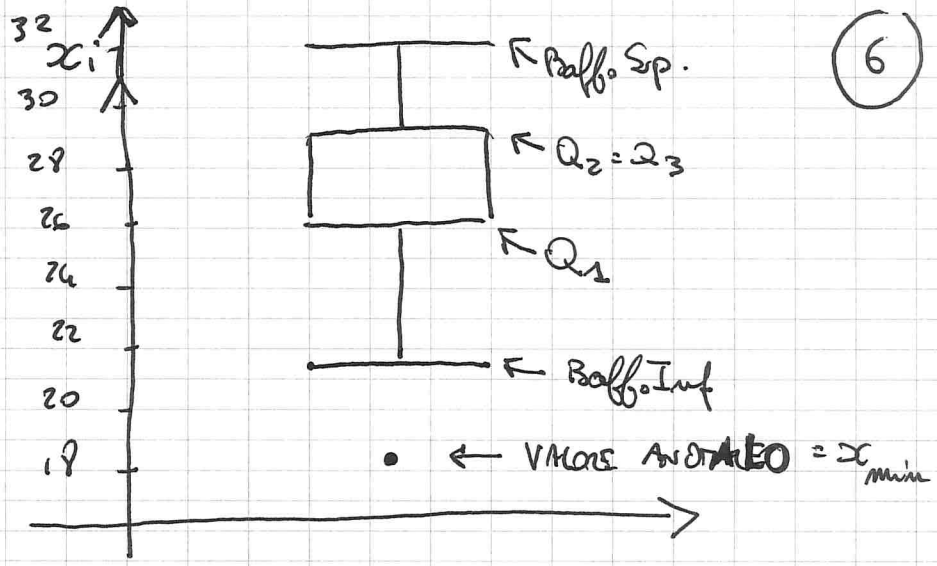
$= 1.7086$

- $\sigma_W^2 = \sum_{j=1}^{k} \sigma^2_{Y|x_j} \cdot \dfrac{n_j}{n} = 31.5 \cdot \dfrac{30}{80} + 36.36 \cdot \dfrac{50}{80} = 34.5375$

$\Rightarrow \sigma_Y^2 = Var(Y) \overset{\text{TEOR. SCOM. VAR.}}{=} \sigma_B^2 + \sigma_W^2 = 1.7086 + 34.5375 = 36.2461$

ESERCIZIO 4

x_i	n_i	f_i	F_i	
18	4	0.02	0.02	
23	40	0.2	0.22	
26	36	0.18	0.4	←
29	70	0.35	0.75	⇐
32	50	0.25	1	
	200	1		



- Mediana(X) = $Q_2 = 29$
  $Q_1 = 26$
  $Q_3 = 29$
  $\Rightarrow Q_3 - Q_1 = 29 - 26 = 3$

- $x_{min} = 18$; $x_{max} = 32$

- Baffo Inf. $= \max\{Q_1 - 1.5(Q_3 - Q_1);\ x_{min}\} = \max\{26 - 1.5 \cdot 3;\ 18\}$
  $= \max\{21.5;\ 18\} = 21.5$

  Baffo Sup. $= \min\{Q_3 + 1.5(Q_3 - Q_1);\ x_{max}\} = \min\{29 + 1.5 \cdot 3;\ 32\}$
  $= \min\{33.5;\ 32\} = 32$

ESERCIZIO 5

X	1	a	3
f_i	0.2	b	c

1. $a = 2$

$c = 0.2$, $b = 0.6$ $\Rightarrow$

X	1	2	3
f_i	0.2	0.6	0.2

2. $M(X) = 2$ ($M(X)$ = centro se distr. simm.)

$= \text{Mediana}(X)$

$M(X^2) = 1^2 \cdot 0.2 + 2^2 \cdot 0.6 + 3^2 \cdot 0.2 = 4.40$

$\text{Var}(X) = 4.40^2 - 2^2 = 0.4$

$\gamma_1 = \dfrac{M((X-\mu)^3)}{\sigma^3} = 0$ PERCHÉ X HA DISTRIB. SIMMETRICA

$X-\mu$	-1	0	1
f_i	0.2	0.6	0.2

$\Rightarrow M((X-\mu)^3) = -1^3 \cdot 0.2 + 0^3 \cdot 0.6 + 1^3 \cdot 0.2 = 0$

3. $Y = a + bX \Rightarrow \gamma_n(Y) = \text{sign}(b) \cdot \gamma_1(X)$

$$\text{sign}(b) = \begin{cases} -1, & \text{se } b < 0 \\ 1, & \text{se } b > 0 \end{cases}$$

$Y = 4 + 2X \Rightarrow$

$M(Y) = 8$

$\text{Var}(Y) = 1.6$

$\gamma_n(Y) = \text{sign}(2) \cdot \gamma_n(X) = 1 \cdot 0 = 0$

$Z = 3X \Rightarrow$

$M(Z) = 6$

$\text{Var}(Z) = 3.6$

$\gamma_n(Z) = \text{sign}(3) \cdot \gamma_n(X) = 1 \cdot 0 = 0.$