

Progetto di Statistica Descrittiva

Roberto Piscopo 512109906 AA 2021/2022

Con il termine *statistica descrittiva* si intende un insieme di tecniche e strumenti finalizzati ad assolvere uno dei principali compiti assegnati della Statistica: descrivere, rappresentare e sintetizzare in maniera opportuna un insieme o campione di dati relativamente ad un problema (popolazione) di interesse.

Per *popolazione* si intende la totalità dei casi, ovvero delle unità sulle quali è possibile rilevare una variabile di interesse.

Popolazione di interesse: Punti per Partita Regular Season NBA 2021/2022 di 581 giocatori.

Per *campione* si intende invece un insieme finito di n unità che si può ritenere rappresentativo dell'intera popolazione (che per definizione ha invece numerosità infinita).

Campione: Punti per partita regular season NBA 2021/2022 di 50 giocatori. ($n=50$)

Nome	P	Nome	P	Nome	P	Nome	P	Nome	P
Joel Embiid	30	Jayson Tatum	26	Bradley Beal	23	Dejounte Murray	21	Terry Rozier	19
LeBron James	30	Devin Booker	26	Anthony Davis	23	Tyler Herro	20	Jerami Grant	19
Giannis Antetokounmpo	29	Donovan Mitchell	25	De'Aaron Fox	23	Klay Thompson	20	Malcolm Brogdon	19
Kevin Durant	29	Stephen Curry	25	Pascal Siakam	22	Fred VanVleet	20	Bam Adebayo	19
Luka Doncic	28	Karl-Anthony Towns	24	Brandon Ingram	22	Kristaps Porzingis	20	Norman Powell	19
Trae Young	28	Shai Gilgeous-Alexander	24	CJ McCollum	22	Miles Bridges	20	Domantas Sabonis	18
DeMar DeRozan	27	Zach LaVine	24	James Harden	22	LaMelo Ball	20	Jordan Poole	18
Kyrie Irving	27	Paul George	24	Darius Garland	21	Julius Randle	20	Russell Westbrook	18
Ja Morant	27	Damian Lillard	24	Jimmy Butler	21	Khris Middleton	20	Dillon Brooks	18
Nikola Jokic	27	Jaylen Brown	23	Anthony Edwards	21	RJ Barrett	20	Gary Trent Jr.	18

Per descrivere e sintetizzare l'informazione campionaria di un fenomeno numerico di interesse, la statistica descrittiva si focalizza su 3 principali aspetti:

1. la descrizione e la forma della distribuzione
2. la posizione o tendenza centrale
3. la variabilità o dispersione

Gli strumenti messi a disposizione dalla statistica descrittiva possono essere sia di tipo grafico sia numerico. In questo ultimo caso si tratta di opportuni indici di sintesi, che in unico valore esprimono una specifica caratteristica della distribuzione dei dati: la tendenza centrale, la variabilità e la forma della distribuzione.

Grafici:

- Tabella frequenze, grafico a linee, grafico a barre, grafico a torta.
- Box-plot

Indici di sintesi:

- Indici di posizione o tendenza centrale
media, mediana, moda.
- Indici di variabilità o dispersione
varianza, deviazione standard, scarto medio assoluto, ampiezza del campo di variazione, coefficiente di variazione.
- Indice di forma
indice di asimmetria, indice di curtosi.

Elenco delle possibili modalità

L'insieme dei valori assumibili:

Modalità: 18,19,20,21,22,23,24,2,26,27,28,29,30 (**k=13**)

Possiamo definire vari tipi di frequenze per ogni modalità.

Con il termine **frequenza assoluta** si intende il numero/conteggio di unità che cadono in una determinata classe (intervallo).

Con il termine **frequenza relativa** si intende la frazione (o la percentuale) di unità (rispetto al numero totale di unità) che cadono in una determinata classe (intervallo).

Se sommiamo via via le frequenze in maniera cumulata rispetto agli intervalli, si ottiene la cosiddetta **frequenza cumulata**, che ci dice quante osservazioni cadono fino ad una certa soglia.

Per costruzione, il valore della frequenza cumulata rispetto all'ultima soglia sarà il numero totale di osservazioni o il valore 100% rispettivamente per la frequenza cumulata assoluta o relativa.

La frequenza cumulata può essere sia assoluta sia relativa.

Nostro campione:

18,18,18,18,18,19,19,19,19,19,20,20,20,20,20,20,20,20,20,20,21,21,21,21,22,22,22,22,23,23,23,23,24,24,24,24,24,25,25,26,26,27,27,27,27,28,28,29,29,30,30

Tabella frequenze, grafico a linee, grafico a barre, grafico a torta

Tabella Frequenze assolute, cumulative + frequenze relative, cumulative

i	v_i	f_i	p_i	F_i	P_i
1	18	5	0.1	5	0.1
2	19	5	0.1	10	0.2
3	20	9	0.18	19	0.38
4	21	4	0.08	23	0.46
5	22	4	0.08	27	0.54
6	23	4	0.08	31	0.62
7	24	5	0.1	36	0.72
8	25	2	0.04	38	0.76
9	26	2	0.04	40	0.8
10	27	4	0.08	44	0.88
11	28	2	0.04	46	0.92
12	29	2	0.04	48	0.96
13	30	2	0.04	50	1.0

Grafico a linee (polygon graph): Sull'asse delle ascisse sono riportate le modalità; sull'asse delle ordinate vi è la frequenza assoluta di ciascun valore, ed i punti sono uniti da una spezzata.

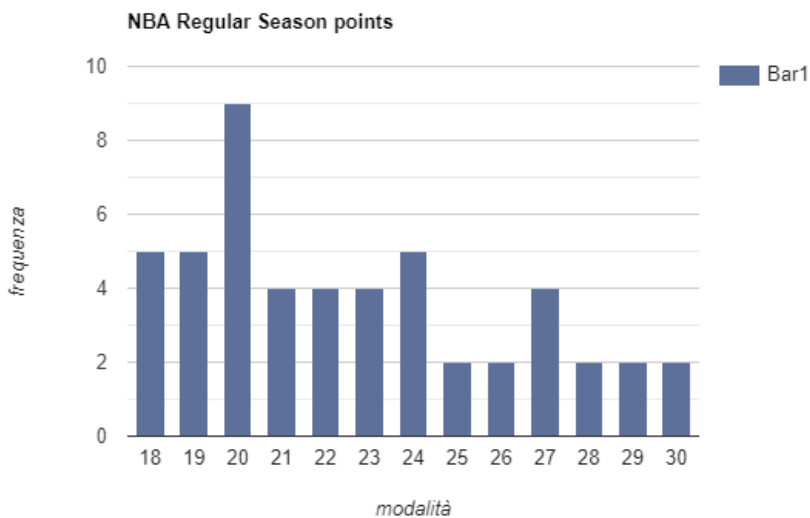
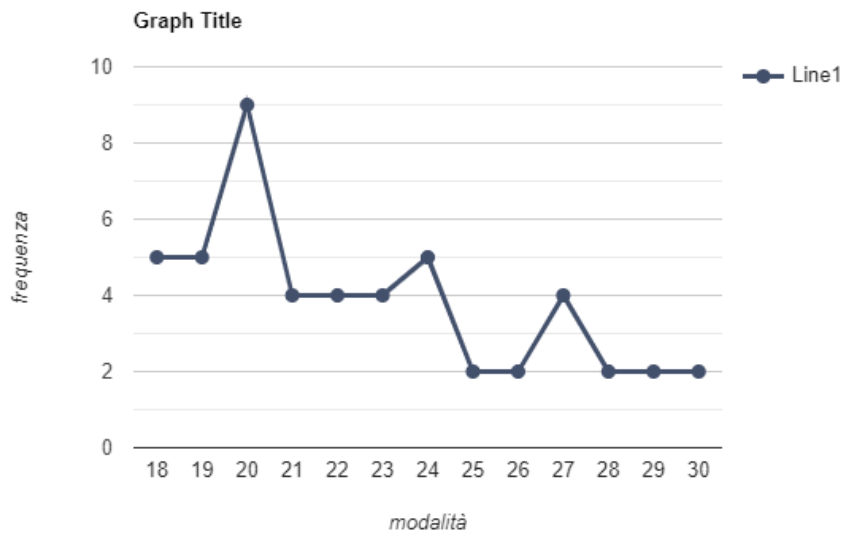
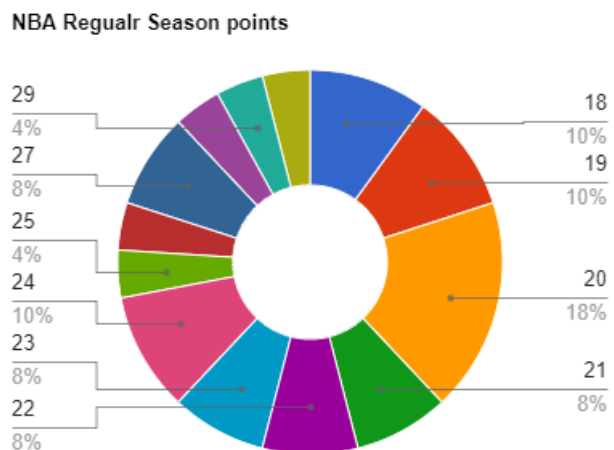


Grafico a barre (bar graph): Sull'asse delle ascisse sono riportate le modalità; sull'asse delle ordinate è riportata la frequenza assoluta di ciascun valore, rappresentata da rettangoli.

Grafico a torta: ogni settore ha angolo al centro proporzionale alla frequenza della modalità corrispondente.



Indici di sintesi

Indici di posizione

Gli indici di posizione (o di tendenza centrale) sono misure che consentono di sintetizzare i dati osservati x_1, x_2, \dots, x_n con un solo valore numerico che sia rappresentativo dei dati stessi.

Gli indici di posizione più adoperati sono tre (la media, la mediana e la moda).

Campione:

18,18,18,18,18,19,19,19,19,19,20,20,20,20,20,20,20,20,20,20,21,21,21,21,22,22,22,22,23,23,23,23,24,24,24,24,24,25,25,26,26,27,27,27,27,28,28,29,29,30,30

$$\text{Media campionaria } \bar{x} = \frac{1}{n} \sum_{i=1}^k v_i f_i = \frac{5}{50} 18 + \frac{5}{50} 19 + \frac{9}{50} 20 + \frac{4}{50} 21 + \frac{4}{50} 22 + \frac{4}{50} 23 + \frac{5}{50} 24 + \frac{2}{50} 25 + \frac{2}{50} 26 + \frac{4}{50} 27 + \frac{2}{50} 28 + \frac{2}{50} 29 + \frac{2}{50} 30 = 22,66$$

Mediana Campionaria (media aritmetica tra i due elementi centrali essendo n pari) = media aritmetica dei valori in posizione 25 e 26 in ordine crescente:

$$\tilde{x} = \frac{[x_{(25)} + x_{(26)}]}{2} = 22$$

Mediana campionaria inferiore alla media campionaria 22,66

Moda campionaria: la moda campionaria di un campione di dati, se esiste, è l'unico valore che ha frequenza massima, e si denota con \hat{x} :

(valore frequenza massima) $\hat{x} = 20$

La distribuzione dei dati è unimodale (una sola moda)

Indici di variabilità

Gli indici di variabilità (o di dispersione) sono delle misure che descrivono la variabilità dei dati osservati x_1, x_2, \dots, x_n , e che pertanto consentono di valutare l'informazione fornita dall'indice di posizione adoperato, dando così una descrizione più accurata della struttura dei dati.

Per valutare la variabilità dei dati adoperiamo :

Varianza campionaria: Per stabilire quanto i dati siano o concentrati o dispersi attorno alla media campionaria, consideriamo le distanze dei dati dalla media campionaria, le eleviamo al quadrato e facciamo la media aritmetica

$$s^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2 \rightarrow 12.555510204082$$

Deviazione standard campionaria

Come la varianza campionaria, la deviazione standard è più grande quanto più i dati si discostano dalla media campionaria

$$s = \sqrt{12.555510204082} = 3.5433755381108$$

Scarto medio assoluto: è dato dalla media aritmetica degli scarti assoluti dalla media campionaria:

$$s_a = 2.9928$$

Ampiezza del campo di variazione: differenza tra i valori estremi delle osservazioni

$$W := x_n - x_1 = 30 - 18 = 12$$

Coefficiente di variazione: Indice adimensionale di variabilità dato dal rapporto tra la deviazione standard campionaria e la media campionaria:

$$CV = \frac{s}{\bar{x}} = \frac{3.507763}{22.66} = 0.1548$$

Indice di Forma

Gli indici di forma misurano caratteristiche relative alla forma della distribuzione dei dati.

Indice di asimmetria (Skewness)

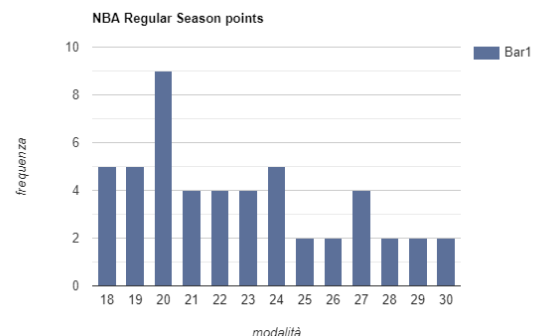
Si usa per stabilire se i dati sono caratterizzati da simmetria centrale:

- $\gamma > 0$: la distribuzione di dati presenta una coda più lunga a destra (asimmetria positiva),
- $\gamma < 0$: la distribuzione presenta una coda più lunga a sinistra (asimmetria negativa).

$$\gamma := \frac{1}{s^3} \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^3. \quad \gamma = 0.527767654$$

$$\begin{aligned} & (18 - 22.66)^3 + (18 - 22.66)^3 + (18 - 22.66)^3 + (18 - 22.66)^3 + (18 - 22.66)^3 + (19 - 22.66)^3 + (19 - 22.66)^3 + (19 - 22.66)^3 + (19 - 22.66)^3 + \\ & (19 - 22.66)^3 + (20 - 22.66)^3 + (20 - 22.66)^3 + (20 - 22.66)^3 + (20 - 22.66)^3 + (20 - 22.66)^3 + (20 - 22.66)^3 + (20 - 22.66)^3 + (20 - 22.66)^3 + \\ & (20 - 22.66)^3 + (21 - 22.66)^3 + (21 - 22.66)^3 + (21 - 22.66)^3 + (21 - 22.66)^3 + (21 - 22.66)^3 + (22 - 22.66)^3 + (22 - 22.66)^3 + (22 - 22.66)^3 + (22 - 22.66)^3 + \\ & (23 - 22.66)^3 + (23 - 22.66)^3 + (23 - 22.66)^3 + (23 - 22.66)^3 + (24 - 22.66)^3 + (24 - 22.66)^3 + (24 - 22.66)^3 + (24 - 22.66)^3 + (24 - 22.66)^3 + \\ & (25 - 22.66)^3 + (25 - 22.66)^3 + (26 - 22.66)^3 + (26 - 22.66)^3 + (27 - 22.66)^3 + (27 - 22.66)^3 + (27 - 22.66)^3 + (27 - 22.66)^3 + (28 - 22.66)^3 + \\ & (28 - 22.66)^3 + (29 - 22.66)^3 + (29 - 22.66)^3 + (30 - 22.66)^3 + (30 - 22.66)^3 * 1/50^{\circ} \quad 1/(3.5434)^{\circ} \approx 0.5277676 \end{aligned}$$

$\gamma > 0$: la distribuzione di dati presenta una coda più lunga a destra (asimmetria positiva)



Indice di curtorsi (o di appiattimento):

E' utile per stabilire se la distribuzione è poco o molto appiattita,

$$k := \left[\frac{1}{s^4} \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^4 \right] - 3$$

$$k = 2.36232649$$

$k > 0$: è presente un eccesso di dati nelle classi centrali (i dati hanno distribuzione leptocurtica)

Percentile

Il percentile k-esimo di un campione di dati è un valore che è maggiore di una percentuale k dei dati, e minore della restante percentuale $100 - k$, dove k è un numero tra 0 e 100.

Esempio 60-esimo percentile:

$$n=50 \quad k=60 \quad p=60/100$$

$$np=50 \cdot 0,6=30$$

$$n(1-p)=50(1-0,6)=20$$

60-esimo percentile è il numero tale che almeno 30 dati sono \leq ad esso, ed almeno 20 dati sono \geq ad esso ---->23

- Il 25-esimo percentile si dice primo quartile. (**Q1**)
- Il 50-esimo percentile si dice mediana campionaria o secondo quartile. (**Q2**)
- Il 75-esimo percentile si dice terzo quartile. (**Q3**)

I quartili

Primo quartile **Q1** (25-esimo percentile):

$$n=50 \quad p=0.25 \rightarrow np=12.5 \text{ (arr 13)} \quad n(1-p)=37.5 \text{ (38)} \rightarrow 20$$

Secondo quartile **Q2** (50-esimo percentile) :

$$\text{mediana campionaria} = 22$$

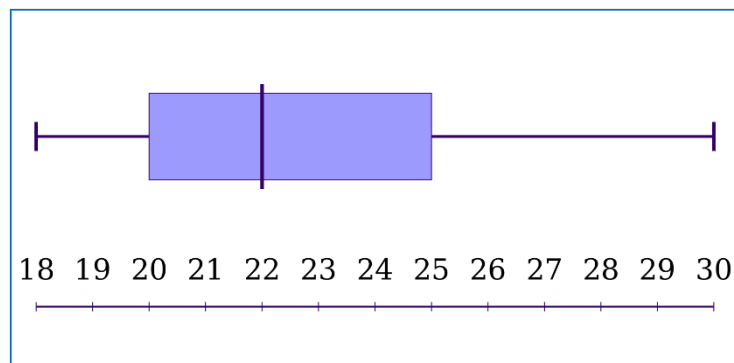
Terzo quartile **Q3** (75-esimo percentile):

$$n=50 \quad p=0.75 \rightarrow np=37.5 \text{ (arr 38)} \quad n(1-p)=12.5 \text{ (13)} \rightarrow 25$$

$$\text{Scarto interquartile } Q3-Q1 = 25-20=5$$

Outliers: nessun dato anomalo nel campione

La rappresentazione grafica dei 5 numeri di sintesi: MIN, Q1, MEDIANA, Q3 e MAX, restituisce il cosiddetto **BOXPLOT**:



Intervalli di previsione

Intervalli del tipo $(\bar{x} - k s, \bar{x} + k s)$, per $k=1,2,3$, con percentuali di dati contenuti negli intervalli.

Spesso risulta utile stabilire quanti dati di un campione sono prossimi alla media campionaria. La disuguaglianza di Chebyshev fornisce una limitazione inferiore per tale numero quando l'ampiezza n del campione è nota.

(Disuguaglianza di Chebyshev) Dato un campione di dati x_1, x_2, \dots, x_n , con media campionaria \bar{x} e deviazione standard campionaria $s > 0$, sia

$$S_k := \{i, 1 \leq i \leq n : |x_i - \bar{x}| < ks\}$$

La disuguaglianza $\frac{|S_k|}{n} > 1 - \frac{1}{k^2}$

mostra che la percentuale di dati x_i compresi in $(\bar{x} - ks, \bar{x} + ks)$ è almeno pari a $1 - \frac{1}{k^2}$.

$$\begin{aligned} 1 - \frac{1}{1} &= 0\% \rightarrow 1 - \frac{1}{1.5^2} = 55,56\% \\ 1 - \frac{1}{4} &= 75\% \\ 1 - \frac{1}{9} &= 88,89\% \end{aligned}$$

Possiamo usare vari tipi di intervalli come rappresentativi di un campione di dati:

- facendo uso del primo e terzo quartile:

$$[Q1 - k(Q3 - Q1), Q3 + k(Q3 - Q1)] \quad k\{1.5, 2, 3\}$$

$$I_1 = [20 - 1.5(25 - 20), 25 + 1.5(25 - 20)] = [12.5, 32.5]$$

$$I_2 = [20 - 2(25 - 20), 25 + 2(25 - 20)] = [10, 35]$$

$$I_3 = [20 - 3(25 - 20), 25 + 3(25 - 20)] = [5, 40]$$

- facendo uso della media campionaria e della deviazione standard campionaria:

$$[\bar{x} - ks, \bar{x} + ks]$$

$$I_1 = [22.66 - 1.5(3.5433), 22.66 + 1.5(3.5433)] = [17.34, 27.97] \quad 88\% \text{ dei dati}$$

$$I_2 = [22.66 - 2(3.5433), 22.66 + 2(3.5433)] = [15.57, 29.74] \quad 98\% \text{ dei dati}$$

$$I_3 = [22.66 - 3(3.5433), 22.66 + 3(3.5433)] = [12.03, 33.28] \quad 100\% \text{ dei dati}$$

Popolazione di interesse: Punti per Partita Regular Season NBA 2021/2022 di 581 giocatori.

Per *campione* si intende invece un insieme finito di n unità che si può ritenere rappresentativo dell'intera popolazione (che per definizione ha invece numerosità infinita).

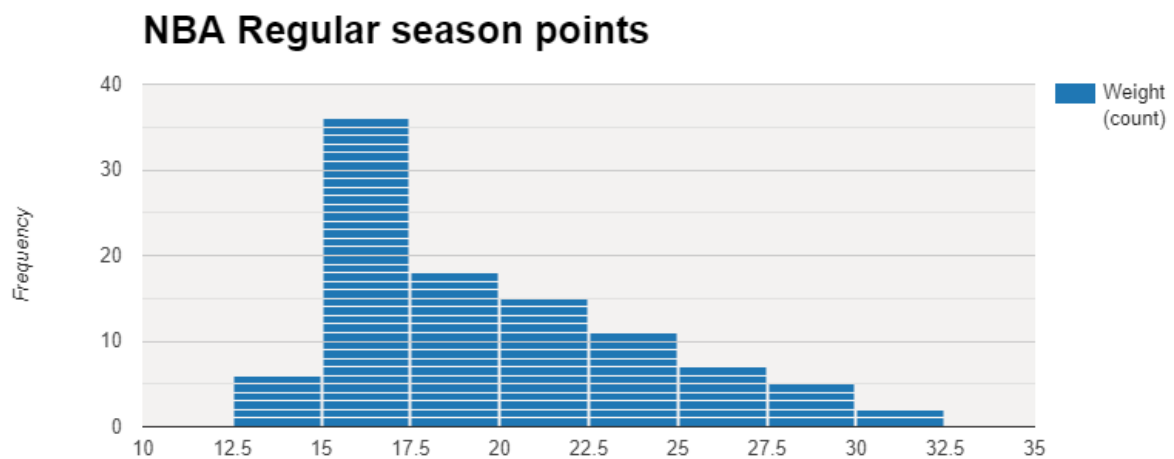
Campione: Media esatta punti per partita regular season NBA 2021/2022 di 100 giocatori. ($n=100$)

Nome	P	Nome	P	Nome	P	Nome	P	Nome	P
Joel Embiid	30.6	Jayson Tatum	26.9	Bradley Beal	23.2	Dejounte Murray	21.1	Terry Rozier	19.3
LeBron James	30.3	Devin Booker	26.8	Anthony Davis	23.2	Tyler Herro	20.7	Jerami Grant	19.2
Giannis Antetokounmpo	29.9	Donovan Mitchell	25.9	De'Aaron Fox	23.2	Klay Thompson	20.4	Malcolm Brogdon	19.1
Kevin Durant	29.9	Stephen Curry	25.5	Pascal Siakam	22.8	Fred VanVleet	20.3	Bam Adebayo	19.1
Luka Doncic	28.4	Karl-Anthony Towns	24.6	Brandon Ingram	22.7	Kristaps Porzingis	20.2	Norman Powell	19.0
Trae Young	28.4	Shai Gilgeous-Alexander	24.5	CJ McCollum	22.1	Miles Bridges	20.2	Domantas Sabonis	18.9
DeMar DeRozan	27.9	Zach LaVine	24.4	James Harden	22.0	LaMelo Ball	20.1	Jordan Poole	18.5
Kyrie Irving	27.4	Paul George	24.3	Darius Garland	21.7	Julius Randle	20.1	Russell Westbrook	18.5
Ja Morant	27.4	Damian Lillard	24.0	Jimmy Butler	21.4	Khris Middleton	20.1	Dillon Brooks	18.4
Nikola Jokic	27.1	Jaylen Brown	23.6	Anthony Edwards	21.3	RJ Barrett	20.0	Gary Trent Jr.	18.3
Jrue Holiday	18.3	Jalen Green	17.3	Harrison Barnes	16.4	Buddy Hield	15.6	Aaron Gordon	15.0
Desmond Bane	18.2	Tobias Harris	17.2	Cole Anthony	16.3	Rudy Gobert	15.6	Seth Curry	15.0
D'Angelo Russell	18.1	Andrew Wiggins	17.2	Jaren Jackson Jr.	16.3	Kevin Porter Jr.	15.6	Evan Mobley	15.0
Bojan Bogdanovic	18.1	Luguentz Dort	17.2	Jalen Brunson	16.3	Marcus Morris Sr.	15.4	Kelly Oubre Jr.	15.0
Christian Wood	17.9	Deandre Ayton	17.2	John Collins	16.2	Tyrese Haliburton	15.3	Josh Hart	14.9
Jonas Valanciunas	17.8	OG Anunoby	17.1	Jarrett Allen	16.1	Scottie Barnes	15.3	Lauri Markkanen	14.8
Nikola Vucevic	17.6	Kyle Kuzma	17.1	Saddiq Bey	16.1	Franz Wagner	15.2	Chris Paul	14.7
Tyrese Maxey	17.5	Keldon Johnson	17.0	Jordan Clarkson	16.0	Bogdan Bogdanovic	15.1	Jaylen Hoard	14.7
Cade Cunningham	17.4	Caris LeVert	17.0	Collin Sexton	16.0	Jusuf Nurkic	15.0	Will Barton	14.7
Anfernee Simons	17.3	Reggie Jackson	16.8	Gordon Hayward	15.9	Wendell Carter Jr.	15.0	Bobby Portis	14.6

Numero di modalità elevato -> dati raccolti in gruppi o classi.

Il grafico a barre delle frequenze delle classi prende il nome di **istogramma**

Classi di ampiezza 2.5



F1	F2	F3	F4	F5	F6	F7	F8
6	36	18	15	11	7	5	2

Indici di posizione

Media campionaria $\bar{x} = 19.653$

Mediana Campionaria: media aritmetica dei valori in posizione 50 e 51 in ordine crescente:

$$\tilde{x} = \frac{[x_{(50)} + x_{(51)}]}{2} = 18.3$$

Mediana campionaria inferiore alla media campionaria

Moda campionaria (valore frequenza massima) $\hat{x} = 15$

La distribuzione dei dati è unimodale (una sola moda)

Indici di variabilità

Varianza campionaria:

$$s^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2 \rightarrow 18.9677687$$

Deviazione standard campionaria

$$s = \sqrt{18.9677687} = 4.35520019$$

Scarto medio assoluto:

$$s_a = 3.5696$$

Ampiezza del campo di variazione:

$$w := x_n - x_1 = 30.6 - 14.6 = 16$$

Coefficiente di variazione:

$$cv = \frac{s}{\bar{x}} = 0.221604854$$

Indice di Forma

Indice di asimmetria (Skewness):

$$\gamma = 0.913047$$

$\gamma > 0$: la distribuzione di dati presenta una coda più lunga a destra (asimmetria positiva)

Indice di curtosi

$$k = 2.93873242$$

$k > 0$: è presente un eccesso di dati nelle classi centrali (i dati hanno distribuzione leptocurtica)

Quartili

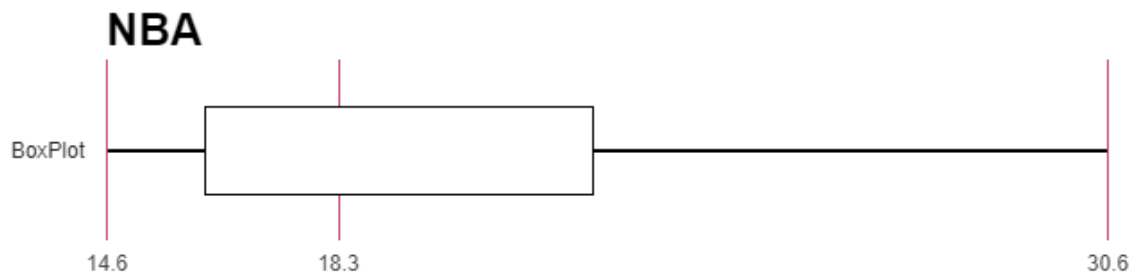
Primo quartile **Q1**: 16.15

Secondo quartile **Q2**: 18.3

Terzo quartile **Q3**: 22.4

Scarto interquartile $Q3 - Q1 = 6.25$

BOXPLOT



Intervallo facendo uso del primo e terzo quartile:

$$[Q1 - k(Q3 - Q1), Q3 + k(Q3 - Q1)] \quad k\{1,2,3\}$$

$$I_1 = [16.15 - 1(6.25), 22.4 + 1(6.25)] = [9.9, 28.65]$$

$$I_2 = [16.15 - 2(6.25), 22.4 + 2(6.25)] = [3.65, 34.9]$$

$$I_3 = [16.15 - 3(6.25), 22.4 + 3(6.25)] = [2.6, 41.25]$$

Dati Bivariati e coefficiente di correlazione campionario

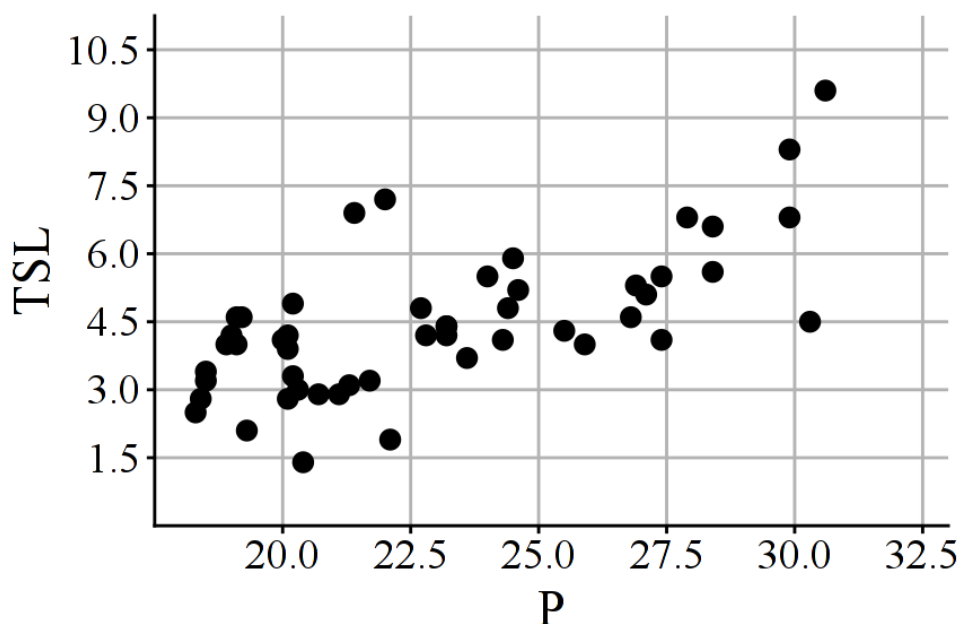
P	TLS		P	TSL
30.6	9.6		22.1	1.9
30.3	4.5		22.0	7.2
29.9	8.3		21.7	3.2
29.9	6.8		21.4	6.9
28.4	5.6		21.3	3.1
28.4	6.6		21.1	2.9
27.9	6.8		20.7	2.9
27.4	4.1		20.4	1.4
27.4	5.5		20.3	3.0
27.1	5.1		20.2	4.9
26.9	5.3		20.2	3.3
26.8	4.6		20.1	2.8
25.9	4.0		20.1	4.2
25.5	4.3		20.1	3.9
24.6	5.2		20.0	4.1
24.5	5.9		19.3	2.1
24.4	4.8		19.2	4.6
24.3	4.1		19.1	4.0
24.0	5.5		19.1	4.6
23.6	3.7		19.0	4.2
23.2	4.2		18.9	4.0
23.2	4.4		18.5	3.2
23.2	4.4		18.5	3.4
22.8	4.2		18.4	2.8
22.7	4.8		18.3	2.5

Campione n: 50

X= Punti per Partita

Y= Tiri liberi Segnati

Uno strumento utile a visualizzare campioni bivariati è il **diagramma a dispersione**, che consiste nella rappresentazioni sul piano cartesiano di tanti punti per quante sono le osservazioni (x_i , y_i), ognuno tracciato nelle corrispondenti coordinate.



Coefficiente di correlazione

Il coefficiente di correlazione è una misura specifica usata nell'analisi della correlazione per quantificare la forza della relazione lineare tra due variabili. Nei report, tale coefficiente è indicato con la lettera r .

- Si parla di correlazione positiva quando a valori elevati di x corrispondono valori elevati di y , e a valori bassi di x corrispondono valori bassi di y . ($r > 0$)
- Viceversa, si parla di correlazione negativa quando a valori elevati di x corrispondono valori bassi di y , e a valori bassi di x corrispondono valori elevati di y . ($r < 0$)

Una risposta grossolana alla questione della correlazione di può ottenere osservando il diagramma a dispersione.

Valori X

$$\sum = 223.4$$

Media Campionaria $\bar{x} = 23.058$

$$\sum (x_i - \bar{x})^2 = SSx = 655.262$$

Valori Y

$$\sum = 1152.9$$

Media Campionaria $\bar{y} = 4.468$

$$\sum (y_i - \bar{y})^2 = SSy = 125.389$$

X e Y

$N = 50$

$\sum (x_i - \bar{x})(y_i - \bar{y}) = 190.663$ --> l'intero campione di dati bivariati mostra una correlazione positiva

Calcolo coefficiente di correlazione

$$r := \frac{1}{(n-1)s_x s_y} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})$$
$$= \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \sum_{i=1}^n (y_i - \bar{y})^2}}$$
$$r = \frac{190.663}{\sqrt{((655.262)(125.389))}} = 0.6652$$

Il valore assoluto di r è una misura della forza della correlazione esistente tra i valori x_i e y_i .

$$|r| = 0.6652$$

Questa è una correlazione positiva abbastanza forte, il che significa che c'è una tendenza a valori X alti con valori Y alti (e viceversa).

Spiegazioni plausibile:

La presenza di un punteggio elevato per partita indica la capacità e bravura di un giocatore, di conseguenza, è probabile che molti punti segnati per partita includano anche i tiri liberi.