Journal homepage http://revistas.unitru.edu.pe/index.php/SSMM

SELECCIONES MATEMÁTICAS
Universidad Nacional de Trujillo
ISSN: 2411-1783 (Online)
2026; Vol.13(1):85-110.

Statistical analysis of the Messi vs Cristiano Ronaldo rivalry

Análisis estadístico de la rivalidad Messi vs Cristiano Ronaldo

Mississippi Valenzuela, Ingrid D. Castro, Wendy G. Hernández, Isabela Solis, María C. Guerra, Wendy N. Del Rio, Beatriz A. Rios and Maximiliano Martínez

Received, Mar. 31, 2025; Accepted, Feb. 20, 2026; Published, Jul. 27, 2026

How to cite this article:
Valenzuela M. et al. *Statistical analysis of the Messi vs Cristiano Ronaldo rivalry*. Selecciones Matemáticas. 2026;13(1):85–110. https://doi.org/10.17268/sel.mat.2026.01.07

Abstract

We will put into practice the main methods, tools, and resources for classifying statistical data and graphical representations. We will analyze a data set, how it is defined, its division, and the graphs we can use as a resource for problem-solving. We also reviewed the importance of measures of central tendency, which are used to organize statistical data presented in the various applications of probability and statistics. We apply each measure of central tendency in this work, including the arithmetic mean, harmonic mean, geometric mean, median, mode, and midrange. In addition, we study the variation presented by the data associated with a statistical phenomenon, whose quantification is carried out using dispersion measures, such as range, percentiles, variance and standard deviation in the problem presented here.

Keywords . Stastistic, measures of central tendency, measures of dispersion.

Resumen

Pondremos en práctica los principales métodos, herramientas y recursos para la clasificación de los datos estadísticos y maneras gráficas de representarlos. Analizamos un conjunto de datos, como se definen, su división y la gráfica que podemos utilizar como recurso para resolver problemas. También revisamos la importancia de las medidas de tendencia central, las cuales permiten organizar los datos estadísticos que se presentan en las diferentes aplicaciones de la probabilidad y la estadística. Aplicamos cada medida de tendencia central, en el presente trabajo, como media aritmética, media armónica, media geométrica, mediana, moda y rango medio. Además, estudiamos la variación que presentan los datos asociados a un fenómeno estadístico, cuya cuantificación se lleva a cabo mediante las medidas de dispersión, tales como rango, percentiles, varianza y desviación estándar en el problema presentado aquí.

Palabras clave. Estadística, medidas de tendencia central, medidas de dispersión.

*Universidad Autónoma de Zacatecas, Unidad Académica de Ciencia y Tecnología de la Luz y la Materia, Circuito Marie Curie S/N, Parque Científico y Tecnológico QUANTUM Ciudad del Conocimiento, Zacatecas 98160, Zacatecas, México. **Correspondence author** (mvalenzuelalumat@uaz.edu.mx).

†Universidad Sierra Madre, Vialidad Arroyo de la Plata 126, La Ladrillera, Guadalupe, Zacatecas, México. (icastrocontreras7@gmail.com).

‡Universidad Sierra Madre, Vialidad Arroyo de la Plata 126, La Ladrillera, Guadalupe, Zacatecas, México. (Lupeupe060503@gmail.com).

§Universidad Sierra Madre, Vialidad Arroyo de la Plata 126, La Ladrillera, Guadalupe, Zacatecas, México. (isabelasolis86@gmail.com).

¶Universidad Sierra Madre, Vialidad Arroyo de la Plata 126, La Ladrillera, Guadalupe, Zacatecas, México. (Marii.guerrararudales19@gmail.com).

‖Universidad Sierra Madre, Vialidad Arroyo de la Plata 126, La Ladrillera, Guadalupe, Zacatecas, México. (nereidawnd@gmail.com).

**Universidad Sierra Madre, Vialidad Arroyo de la Plata 126, La Ladrillera, Guadalupe, Zacatecas, México. (beatrizalessandrarioslopez@gmail.com).

††Universidad Sierra Madre, Vialidad Arroyo de la Plata 126, La Ladrillera, Guadalupe, Zacatecas, México. (maximilianomartinezher1@gmail.com).

1. Introduction. Here we present a classification of statistical data [1] [2, 3][4, 5, 6, 7] and graphical ways of representing them. The most general classification is: qualitative data and quantitative data. Quantitative data are further divided into discrete and continuous. Scales are also established to measure or label data, both qualitative and quantitative. For the former, the nominal scale is typically used, although occasionally the nominal scale is also used, although this is not exclusive to them, as this scale is also used for qualitative data. Finally, for the latter, there is the interval scale and the so-called ratio or linear scale. It is worth mentioning that, using graphs called cumulative frequency polygons, one can easily observe the amount of data that does not exceed a certain value.

There are various measures of central tendency to analyze the behavior of a data set. These measures seek to find a value that is in the center of the data. The most common of these measures is the arithmetic mean, and it is the one we use most frequently.

Sometimes we are interested in a data set from a collection with an odd number of data points for which there are as many data points less than or equal to it as there are data points greater than or equal to it. This data point is called the median of a set. When the set has an even number of data points, we say $2n$, the median is the average of the data points in the n and $n+1$ groups.

The harmonic mean of a collection of numbers is defined by the average. When the same distance is traveled several times, the harmonic mean of the different speeds at which each trip was made gives the overall average speed. Given a collection of data, the most frequently repeated value is called the mode. The average of the maximum and minimum values in the collection is called the midrange.

Measures of central tendency are useful for analyzing data associated with a statistical phenomenon, but they are not sufficient. It is necessary to know the degree of variation in the data. This variation is quantified through measures of dispersion. The simplest is the difference between the maximum and minimum values of the data, called the range. Its value can be misleading when a particular data point or group of data points differs significantly from the rest. These data are called aberrant.

For a more detailed analysis, the data is ordered from lowest to highest and certain percentiles are determined, which limit groups of data from above. An example of these are quartiles; in this case, three numbers are determined: Q_1, Q_2 y Q_3 with the property that at least 25% of the data is less than or equal to Q_1, at least 50% is less than Q_2 and at least 75% is less than Q_3. The difference $Q_3 - Q_1$ is called the interquartile range and its value is not affected by the presence of aberrant data, as occurs with the range.

The most widely used measure of dispersion is the standard deviation. It tells us how centered the data are, that is, how representative the mean is of the total data. If the standard deviation is small, most of the data are close to the average; if it is large, however, they are more evenly distributed. Finally, we study TChebysheff's theorem, which allows us, given a number $k > 1$, to determine the percentage of data that are less than k times the standard deviation from the average.

2. Measures of central tendency.

Definition 2.1. *Statistics*[1] *is the area of mathematics that studies how to collect, organize, analyze, interpret, and present data for analysis, in order to find relationships and possible dependencies between physical, chemical, economic, and social phenomena.*

2.1. Data types. When we are faced with the analysis of statistical data, we can identify two types of data:

- Qualitative data: represent diverse categories into which a population of people or things can be classified.
- Quantitative data: they give numerical information, such as how many there are, how tall they are, how much they weigh, etc. When dealing with quantitative data, we distinguish two types: (a) Discrete data are obtained by counting the number of members of a population. (b) Continuous data are obtained through a measurement process.

In the case of continuous data, it is always advisable to use frequency tables, grouping the elements into classes. In the case of discrete data, when there are few data, we can leave them alone and use groupings when the number of values they can take is large.

To represent statistical data, we use different measurement scales.

- Nominal: simply labels the classes into which the objects were grouped. A nominal scale is typically used when working with qualitative data.
- Ordinal: These correspond to an ordered scale, in which we are ranking the data categories, so that each category represents a different level of the attribute being rated.

[1]Godofredo Achenwall (1719-1772), prussian economist, invented the word statistics in 1760.

- Interval: This is used when analyzing quantitative data where it makes sense to calculate the difference between values. The scale used does not necessarily start at zero.
- Linear: This is a type of interval scale in which we can determine how many times greater one measurement is than another.

Year	Goals	Assists	G + A	Cumulative goals	Cumulative assists	Cumulative G+A
2002	5	4	9	5	4	9
2003	1	2	3	6	6	12
2004	13	14	27	19	20	39
2005	15	11	26	34	31	65
2006	25	10	35	59	41	100
2007	34	11	45	93	52	145
2008	35	12	47	128	64	192
2009	30	16	46	158	80	238
2010	48	18	66	206	98	304
2011	60	18	78	266	116	382
2012	63	13	76	329	129	458
2013	69	15	84	398	144	542
2014	61	20	81	459	164	623
2015	57	17	74	516	181	697
2016	55	16	71	571	197	768
2017	53	12	65	624	209	833
2018	49	13	62	673	222	895
2019	39	14	53	712	236	948
2020	44	7	51	756	243	999
2021	47	7	54	803	250	1053
2022	16	4	20	819	254	1073
2023	54	4	58	873	258	1131
2024	43	7	50	916	265	1181

Table 2.1: Cristiano Ronaldo's dates by calendar year.

Year	Goals	Assists	G+A	Cumulative goals	Cumulative assists	Cumulative G+A
2005	3	4	7	3	4	7
2006	12	3	15	15	7	22
2007	31	12	43	46	19	65
2008	22	18	40	68	37	105
2009	41	15	56	109	52	161
2010	60	17	77	169	69	238
2011	59	37	96	228	106	334

2012	91	22	113	319	128	447
2013	45	16	61	364	144	508
2014	58	21	79	422	165	587
2015	52	26	78	474	191	665
2016	59	31	90	533	222	755
2017	54	16	70	587	238	825
2018	51	26	77	638	264	902
2019	50	18	68	688	282	970
2020	27	19	46	715	301	1016
2021	43	18	61	758	319	1077
2022	35	30	65	793	349	1142
2023	28	12	40	821	361	1182
2024	29	18	47	850	379	1229

Table 2.2: Messi's dates by calendar year.

Definition 2.2. *A frequency table indicates how many data items fall into each interval. The intervals into which the data are divided are known as classes.*

For example, in this article:

Interval	Frequency of G+A
$(0, 10]$	2
$(10, 20]$	1
$(20, 30]$	2
$(30, 40]$	1
$(40, 50]$	4
$(50, 60]$	4
$(60, 70]$	3
$(70, 80]$	4
$(80, 90]$	2

Table 2.3: Frequency table of Cristiano Ronaldo's goals+assist.

Interval	Frequency of G+A
$(0, 10]$	1
$(10, 20]$	1
$(20, 30]$	0
$(30, 40]$	2

$(40, 50]$	3
$(50, 60]$	1
$(60, 70]$	5
$(70, 80]$	4
$(80, 90]$	1
$(90, 100]$	1
$(100, 110]$	0
$(110, 120]$	1

Table 2.4: Frequency table of Messi's goals+assist.

Definition 2.3. *The cumulative frequency polygon or ogive is the linear graph obtained from a cumulative frequency table.*

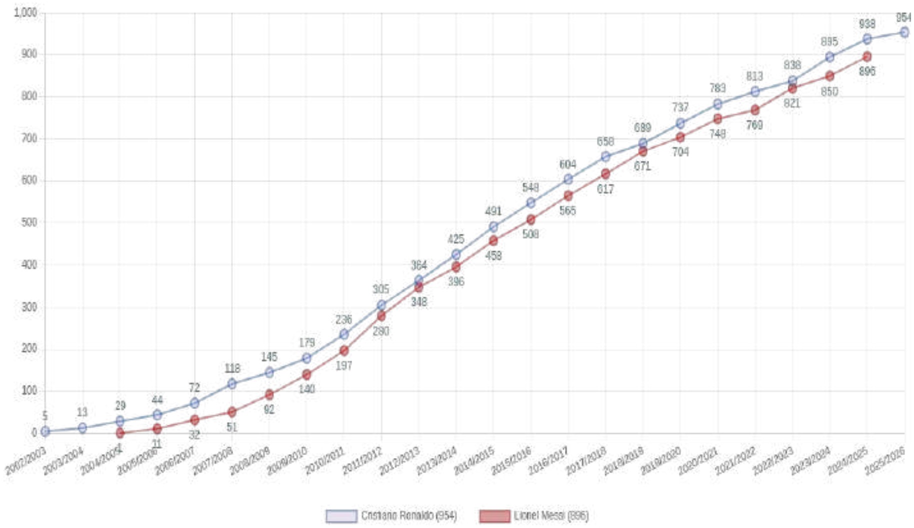

Figure 2.1: Goals accumulated per season.

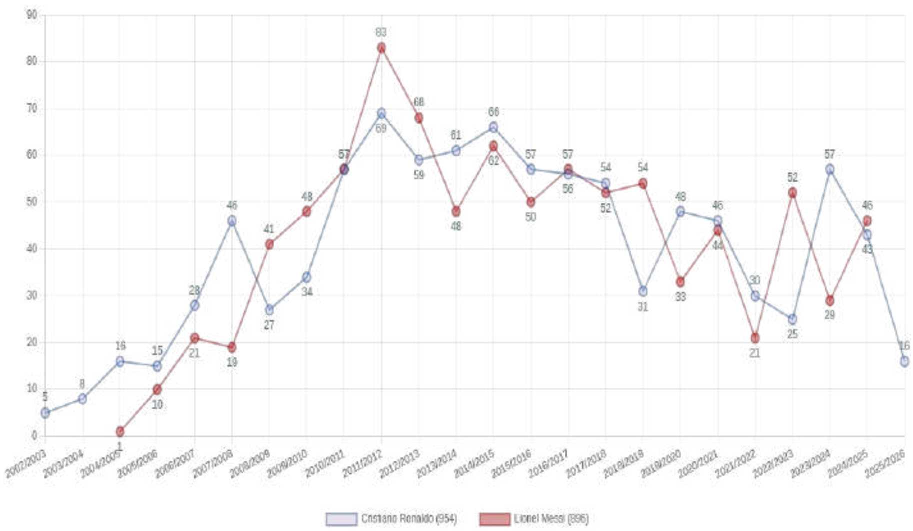

Figure 2.2: Goals per season.

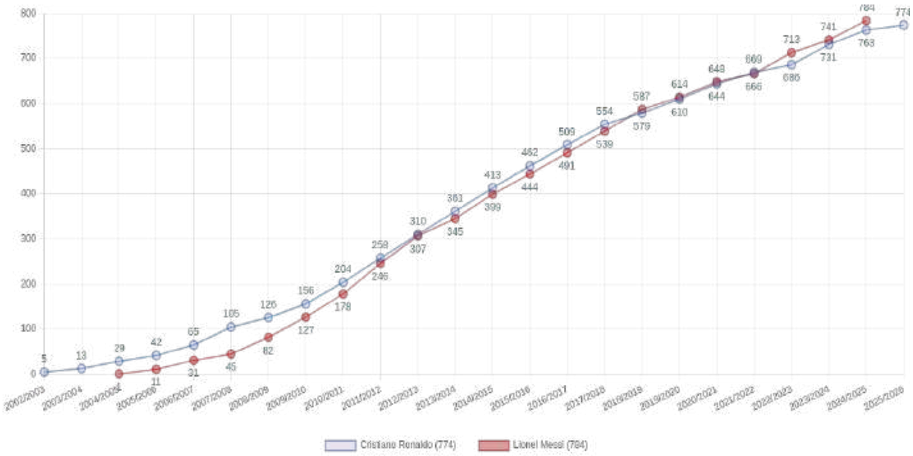

Figure 2.3: Goals accumulated without penalties per season.

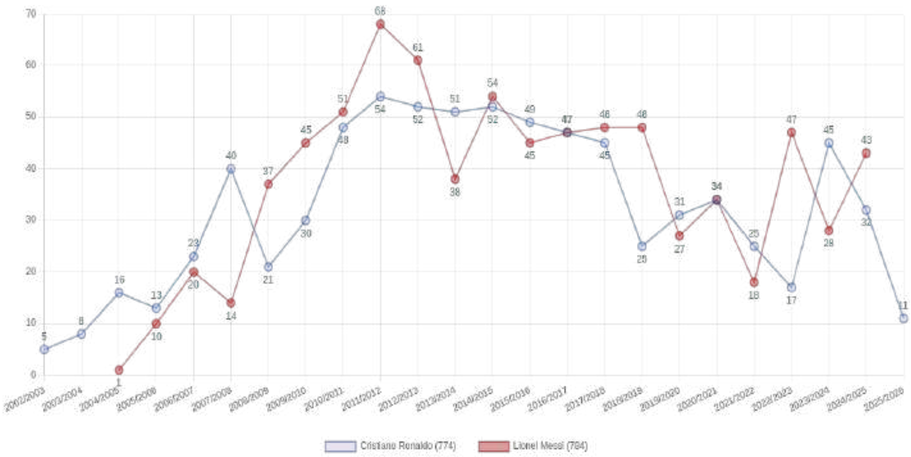

Figure 2.4: Goals (excluding penalties) per season.

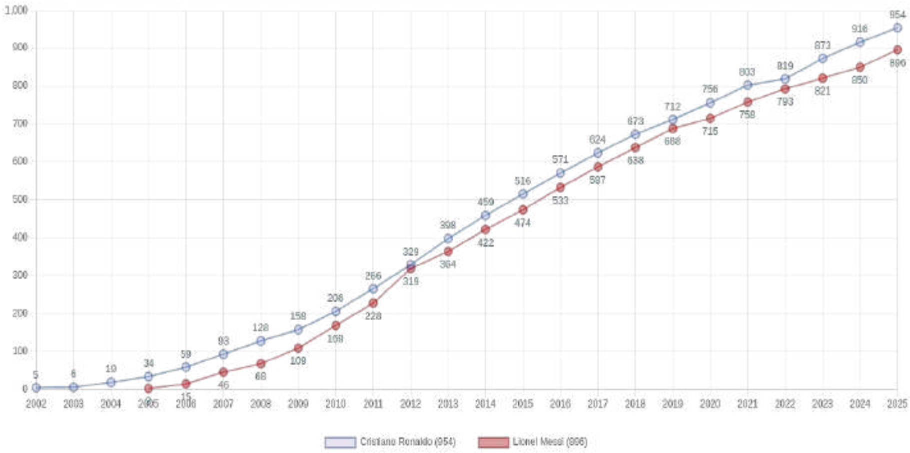

Figure 2.5: Goals accumulated per calendar year.

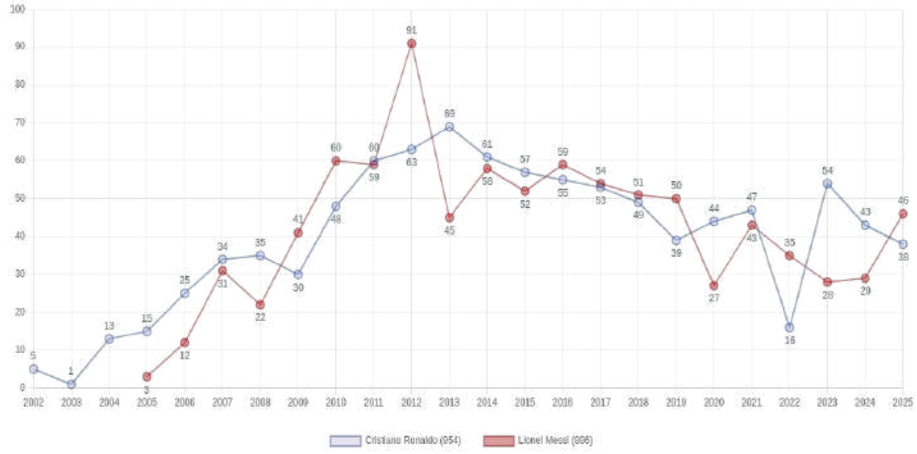

Figure 2.6: Goals per calendar year.

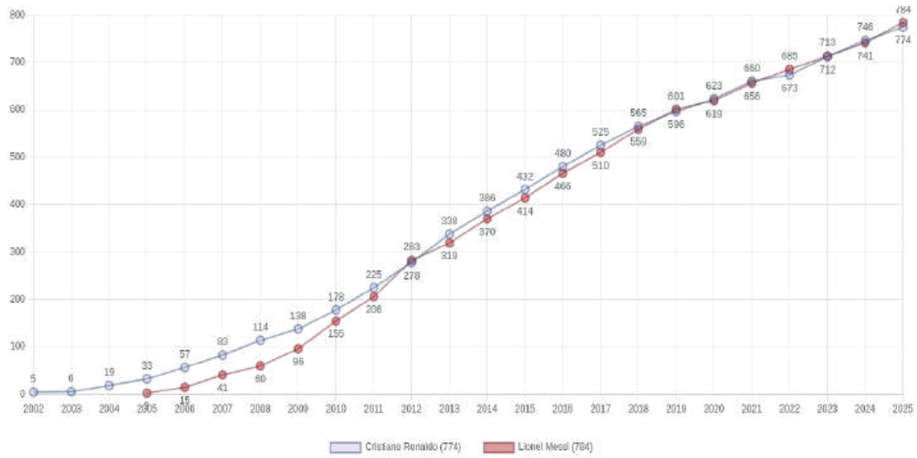

Figure 2.7: Total goals (excluding penalties) per calendar year.

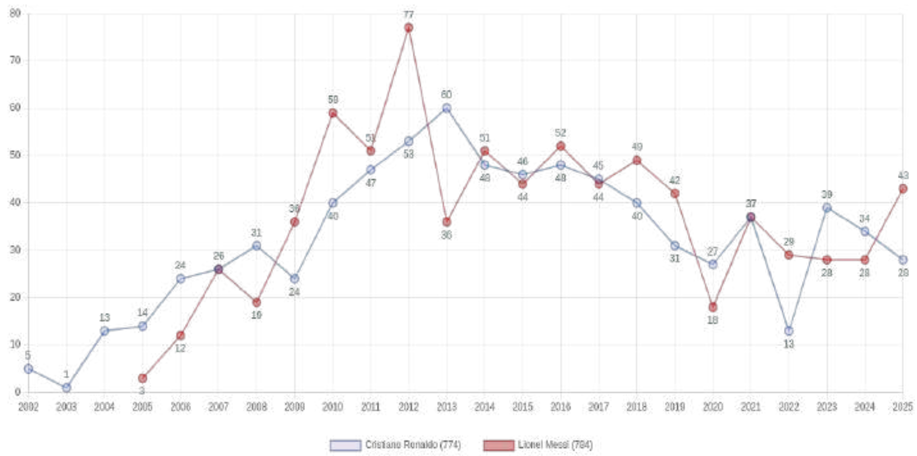

Figure 2.8: Goals (excluding penalties) per calendar year.

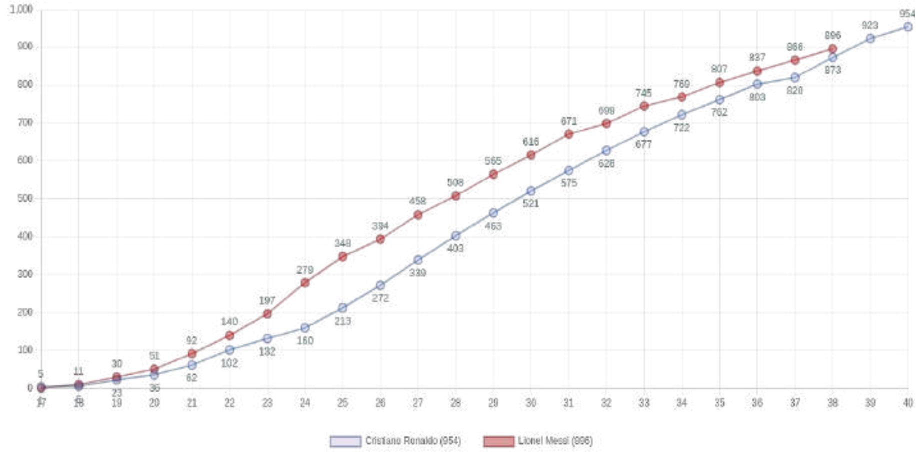

Figure 2.9: Goals accumulated by age.

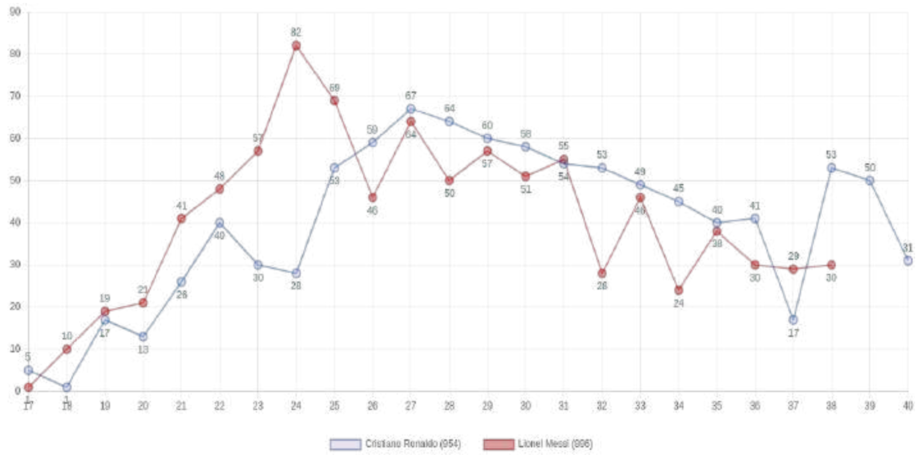

Figure 2.10: Goals by age.

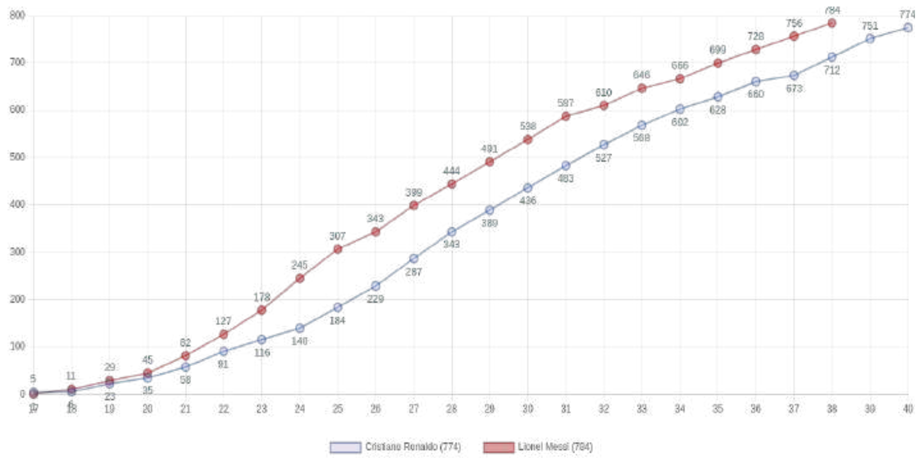

Figure 2.11: Goals (excluding penalties) by age.

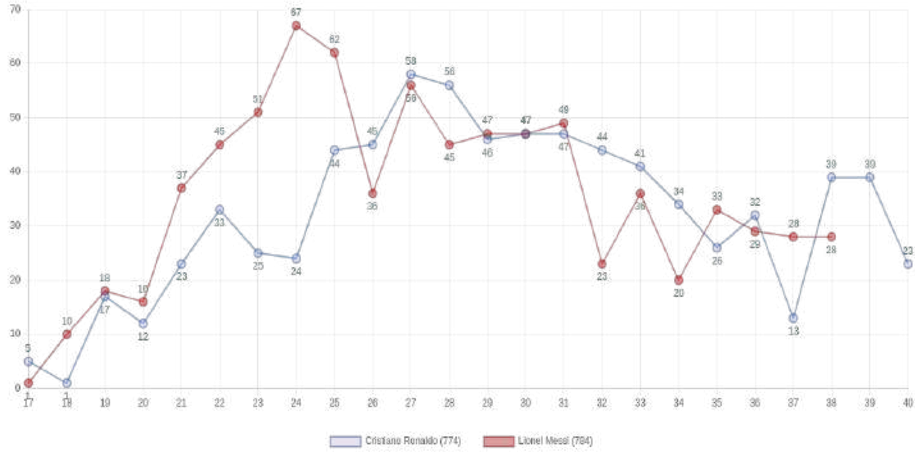

Figure 2.12: Goals (excluding penalties) by age.

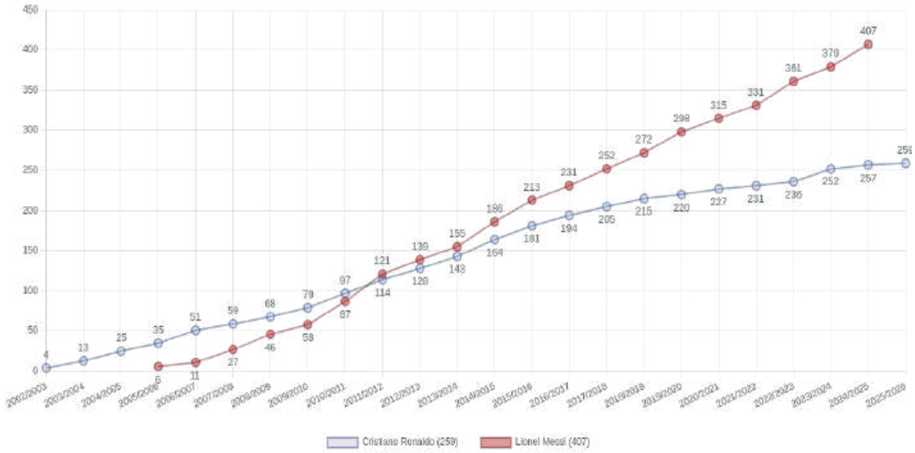

Figure 2.13: Assists (Opta criteria) given by Messi vs Cristiano.

[2]

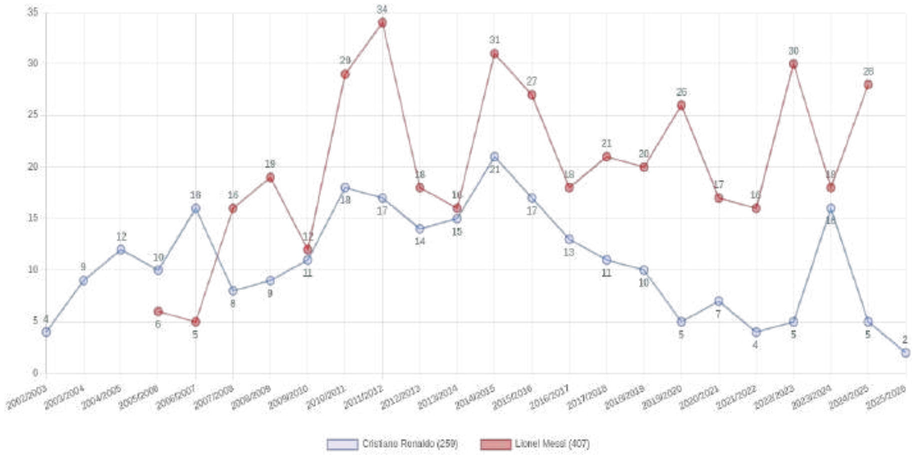

Figure 2.14: Assist per season.

[2] Opta is the most trusted statistics source and comes closest to being the official statistics provider in football. They provide data for the BBC, the Premier League, WhoScored, Squawka, and many other major media outlets. Below is Opta's definition of an assist: The final touch (pass, cross, shot, or any other touch) that leads to the recipient of the ball scoring a goal. If the final touch is deflected by an opposing player, the passer is only awarded an assist if the receiving player would likely have received the ball had the deflection not occurred. Own goals, direct free kicks, direct goals from corners, and penalties do not involve an assist.

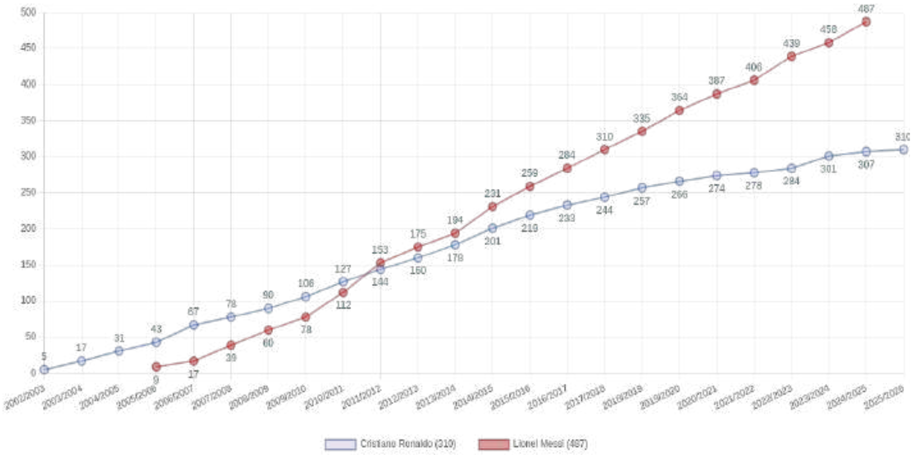

Figure 2.15: Assists (Transfermarket criteria) given by Messi vs Cristiano.

[3]

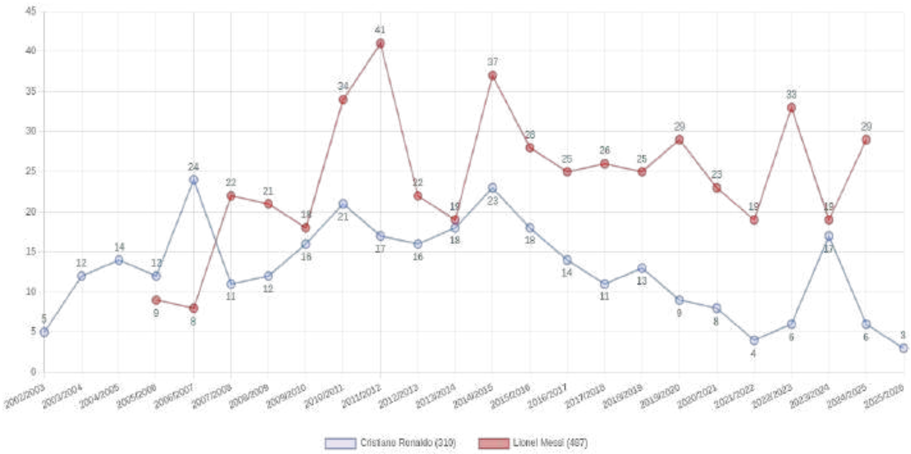

Figure 2.16: Assist per season.

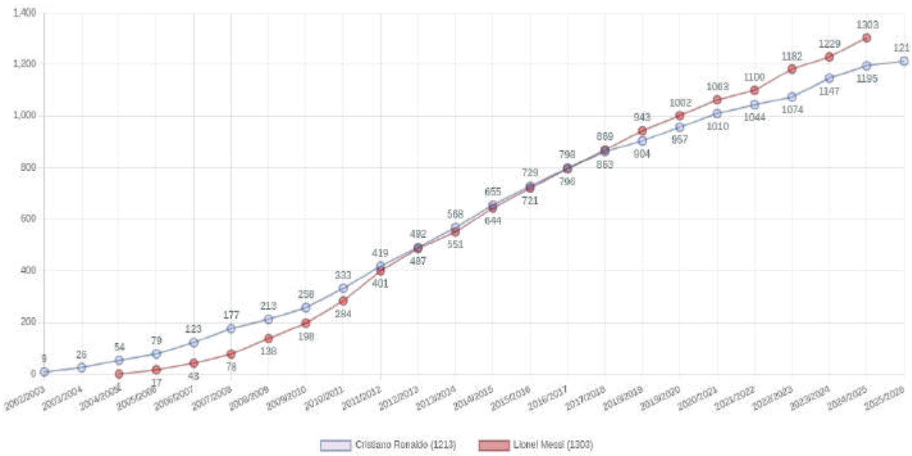

Figure 2.17: Cumulative goal contribution per season.

[3]Transfermarkt's criteria are more flexible in their definition of assists. They include deflected passes, rebounded shots, penalties conceded, and passes or shots that result in an own goal. We use that criteria, although Transfermarkt's figures may differ, as they are quite inconsistent when it comes to compiling assists.

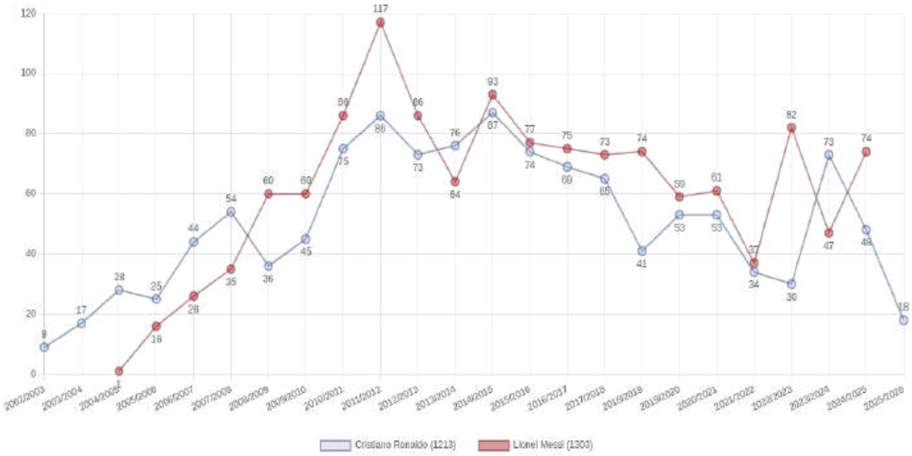

Figure 2.18: Goal contribution per season.

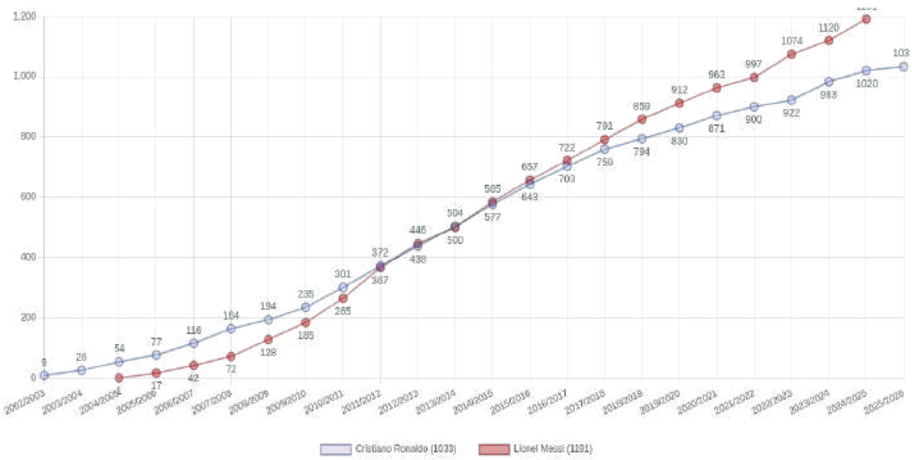

Figure 2.19: Cumulative goal contribution (excluding penalties) per season.

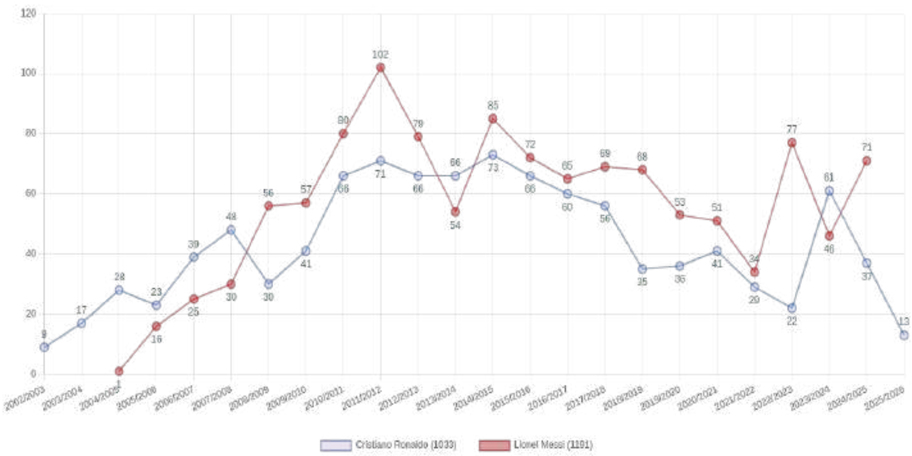

Figure 2.20: Cumulative goal contribution per calendar year.

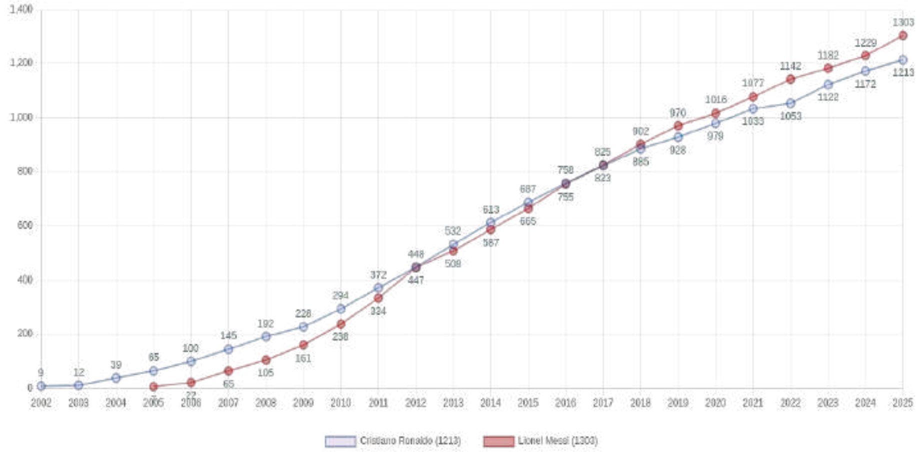

Figure 2.21: Cumulative goal contribution per calendar year.

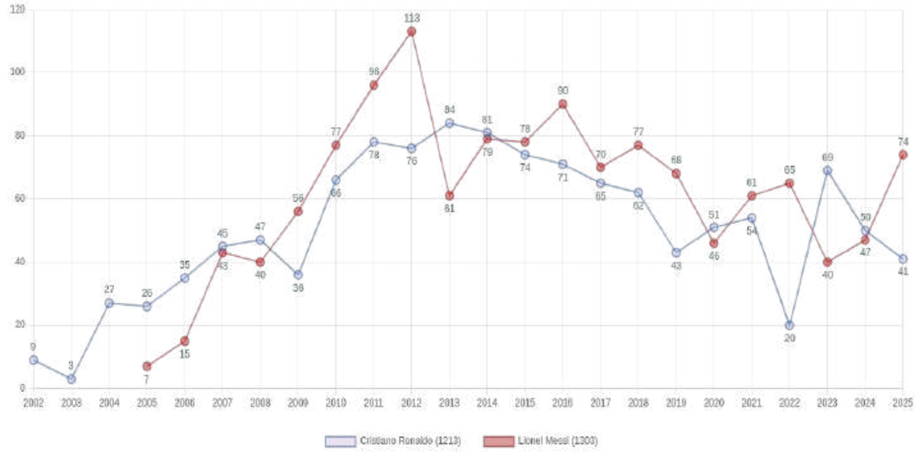

Figure 2.22: Goal contribution per calendar year.

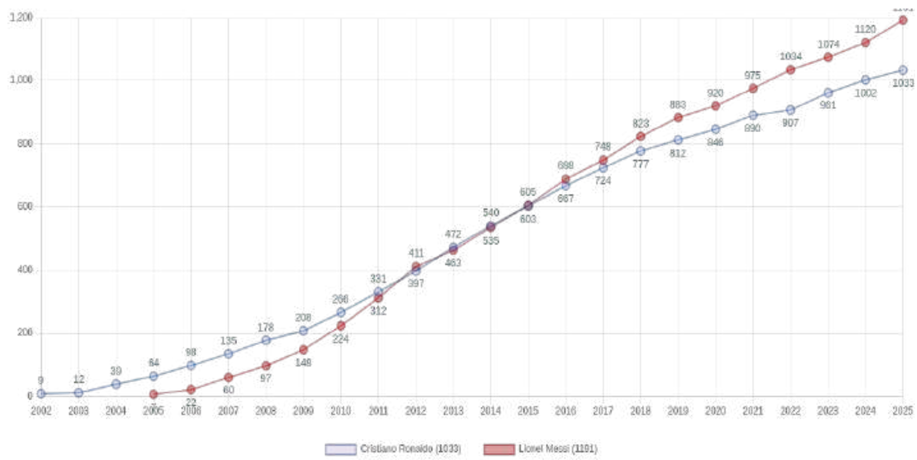

Figure 2.23: Cumulative goal contribution (excluding penalties) per calendar year.

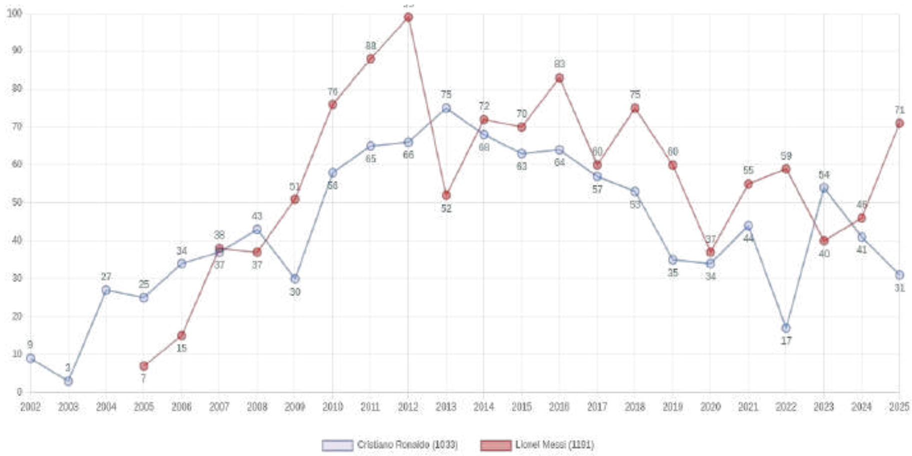

Figure 2.24: Goal contribution (excluding penalties) per calendar year.

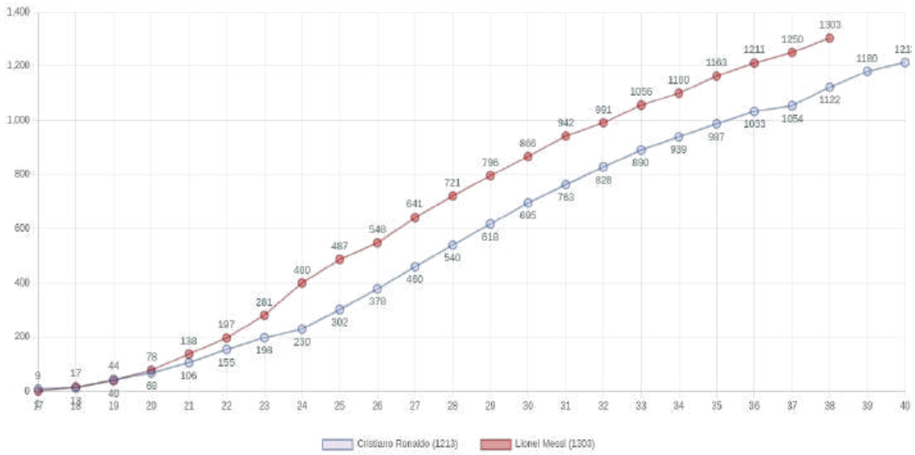

Figure 2.25: Cumulative goal contribution by age.

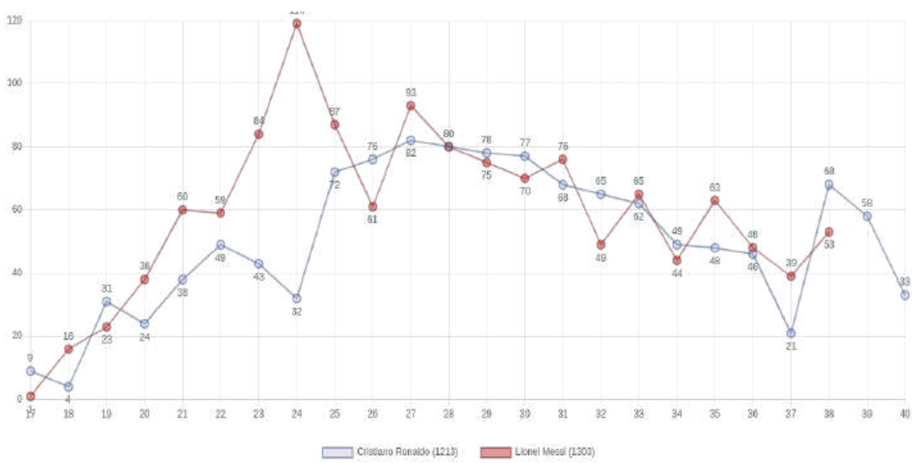

Figure 2.26: Goal contribution by age.

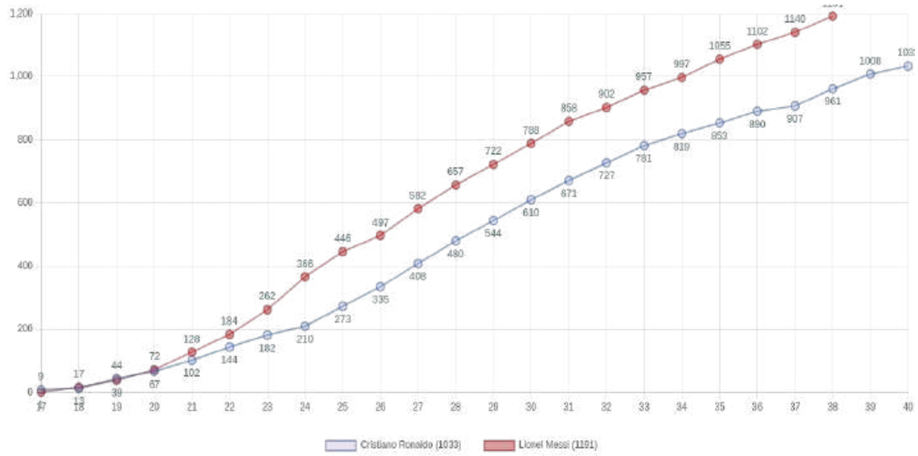

Figure 2.27: Cumulative goal contribution (excluding penalties) by age.

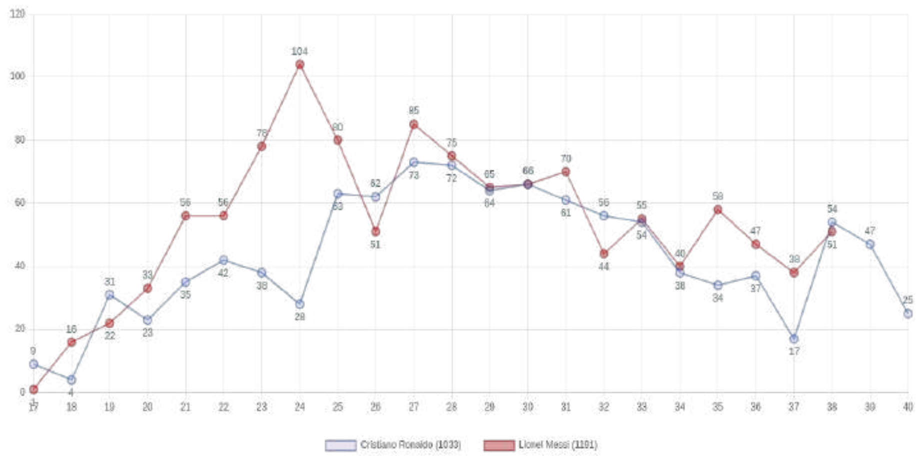

Figure 2.28: Goal contribution (excluding penalties) by age.

Definition 2.4. *The arithmetic mean or average is defined as the sum of all the data divided by the number of data. That is,*

$$\mu = \langle x \rangle = \frac{1}{N} \sum_{k=1}^{N} x_k. \tag{2.1}$$

Now, we compute arithmetic mean or average of goals, assist and generated goals of Cristiano Ronaldo (CR7) and Messi:

$$\langle x \rangle_{CR7}^{Goal} = \frac{1}{23} \sum_{k=1}^{23} x_k = \frac{916}{23} = 39.82, \qquad \langle x \rangle_{Messi}^{Goal} = \frac{1}{20} \sum_{k=1}^{20} x_k = \frac{850}{20} = 42.50, \tag{2.2}$$

$$\langle x \rangle_{CR7}^{Assist} = \frac{1}{23} \sum_{k=1}^{23} x_k = \frac{265}{23} = 11.52, \qquad \langle x \rangle_{Messi}^{Assist} = \frac{1}{20} \sum_{k=1}^{20} x_k = \frac{379}{20} = 18.95, \tag{2.3}$$

$$\langle x \rangle_{CR7}^{G+A} = \frac{1}{23} \sum_{k=1}^{23} x_k = \frac{1181}{23} = 51.34, \qquad \langle x \rangle_{Messi}^{G+A} = \frac{1}{20} \sum_{k=1}^{20} x_k = \frac{1229}{20} = 61.45. \tag{2.4}$$

Definition 2.5. *The geometric mean G of n non-negative numbers $x_1, x_2, ..., x_N$ is defined as*

$$G = \sqrt[N]{x_1 x_2 ... x_N} = \sqrt[N]{\prod_{k=1}^{N} x_k}. \tag{2.5}$$

Then, we compute geometric mean of goals, assist and generated goals of Cristiano Ronaldo (CR7) and Messi:

$$G_{CR7}^{Goals} = \sqrt[23]{\prod_{k=1}^{23} x_k} = \sqrt[23]{1.621181 \times 10^{34}} = 30.72, \tag{2.6}$$

$$G_{CR7}^{Goals} = \sqrt[20]{\prod_{k=1}^{20} x_k} = \sqrt[20]{1.179781 \times 10^{31}} = 35.77, \tag{2.7}$$

$$G_{CR7}^{Assist} = \sqrt[23]{\prod_{k=1}^{23} x_k} = \sqrt[23]{1.071889 \times 10^{23}} = 10.03, \tag{2.8}$$

$$G_{CR7}^{Assist} = \sqrt[20]{\prod_{k=1}^{20} x_k} = \sqrt[20]{2.417914 \times 10^{24}} = 16.56, \tag{2.9}$$

$$G_{CR7}^{G+A} = \sqrt[23]{\prod_{k=1}^{23} x_k} = \sqrt[23]{4.337686 \times 10^{34}} = 35.436, \tag{2.10}$$

$$G_{CR7}^{Assist} = \sqrt[20]{\prod_{k=1}^{20} x_k} = \sqrt[20]{3.591673 \times 10^{34}} = 53.423. \tag{2.11}$$

Definition 2.6. *In general, the harmonic mean a of n numbers is defined as the reciprocal of the average or mean of its reciprocals:*

$$a = \frac{N}{\sum_{k=}^{N} \frac{1}{x_k}}. \tag{2.12}$$

Thence, we have:

$$a_{CR7}^{Goals} = \frac{23}{\sum_{k=}^{23} \frac{1}{x_k}} = \frac{23}{1.784830661} = 12.88, \qquad a_{Messi}^{Goals} = \frac{20}{\sum_{k=}^{20} \frac{1}{x_k}} = \frac{20}{0.856205559} = 23.35,$$
$$a_{CR7}^{Assists} = \frac{23}{\sum_{k=}^{23} \frac{1}{x_k}} = \frac{23}{2.335360881} = 8.11, \qquad a_{Messi}^{Assists} = \frac{20}{\sum_{k=}^{20} \frac{1}{x_k}} = \frac{20}{1.537959092} = 13.01,$$
$$a_{CR7}^{G+A} = \frac{23}{\sum_{k=}^{23} \frac{1}{x_k}} = \frac{23}{0.882480094} = 26.06, \qquad a_{Messi}^{G+A} = \frac{20}{\sum_{k=}^{20} \frac{1}{x_k}} = \frac{20}{0.569312358} = 35.13.$$

Definition 2.7. *Given a data set ordered from smallest to largest, we consider the number in the middle of the data set if the number of data sets is odd, or the average of the two central data sets if it is even. This number is called the median*[4,5,6] *of the data [8, 8, 9, 10], and it has the particularity that there are as many numbers to the left of it as to the right of it.*

We found the medians of CR7 and Messi:

$$\text{Med}_{CR7}^{Goals} = 44, \qquad \text{Med}_{Messi}^{Goals} = \frac{43+45}{2} = 44, \tag{2.13}$$
$$\text{Med}_{CR7}^{Assist} = 12, \qquad \text{Med}_{Messi}^{Assist} = 18, \tag{2.14}$$
$$\text{Med}_{CR7}^{G+A} = 53, \qquad \text{Med}_{Messi}^{G+A} = \frac{61+65}{2} = 63. \tag{2.15}$$

[4] Edward Wright (1561-1615), english mathematician and cartographer, introduced the concept of median in his book "Certain Errors in Navigation", published in 1599.

[5] Antoine Augustin (1801-1877), french mathematician, in 1843 he began to systematically use in his studies the term median, introduced by Wright centuries earlier.

[6] Gustav Theodor Fechner (1801-1877), german philosopher and psychologist, he used the median for the analysis of sociological phenomena.

where we have used the six data sets

$$
\begin{aligned}
\{Goals\}_{CR7} &= \{1,5,13,15,16,25,30,34,35,39,43,44,47,48,49,53,54,55,57,60,61,63,69\}\,,\\
\{Goals\}_{Messi} &= \{3,12,22,27,28,29,31,35,41,43,45,50,51,52,54,58,59,59,60,91\}\,,\\
\{Assists\}_{CR7} &= \{2,4,4,4,4,7,7,10,11,11,12,12,13,13,14,15,16,16,17,18,18,20\}\,,\\
\{Assists\}_{Messi} &= \{3,4,12,12,15,16,16,17,18,18,18,18,19,21,22,26,30,31,37\}\,,\\
\{G+A\}_{CR7} &= \{3,9,20,26,27,35,45,46,47,50,51,53,54,58,62,65,66,71,74,76,78,81,84\}\,,\\
\{G+A\}_{Messi} &= \{7,15,40,40,43,46,47,56,61,61,65,68,70,77,77,78,79,90,96,113\}\,.
\end{aligned}
$$

Definition 2.8. *Given a collection of data, we call mode (Mod) the data that is repeated the most, or the most frequently; these must be repeated more than once and the same number of times.*

The modes of CR7 and Messi are:

$$
\begin{aligned}
\mathrm{Mod}_{CR7}^{Goals} &= 0, & \mathrm{Mod}_{Messi}^{Goals} &= 59,\\
\mathrm{Mod}_{CR7}^{Assist} &= 4, & \mathrm{Mod}_{Messi}^{Assist} &= 18,\\
\mathrm{Mod}_{CR7}^{G+A} &= 0, & \mathrm{Mod}_{Messi}^{G+A} &= \{40,61,77\}\,.
\end{aligned}
$$

Definition 2.9. *The average of the maximum and minimum values of a data set is called the mean range (RanM).*

The Ran of Cristiano Ronaldo and Messi are:

$$
\begin{aligned}
\mathrm{RanM}_{CR7}^{Goals} &= \frac{69+1}{2} = 35, & \mathrm{RanM}_{Messi}^{Goals} &= \frac{91+3}{2} = 47,\\
\mathrm{RanM}_{CR7}^{Assist} &= \frac{20+2}{2} = 11, & \mathrm{RanM}_{Messi}^{Assist} &= \frac{37+3}{2} = 20,\\
\mathrm{RanM}_{CR7}^{G+A} &= \frac{84+3}{2} = 43.5, & \mathrm{RanM}_{Messi}^{G+A} &= \frac{113+7}{2} = 60.
\end{aligned}
$$

3. Measures of dispersion. **Definition 3.1.** *The measures that tell us how much the data vary from its mean value are called measures of dispersion or variability.*

The simplest measure of variability is the range, which is the difference between the maximum and minimum data within a group of data.

Several statistical programs (for example, R[7] [11, 12]) define the range as the interval whose endpoints are the minimum (Min) and maximum (Max) values of the data. This provides more information than simply subtracting the minimum from the maximum.

Definition 3.2. *The simplest variability is the range (Ran), which is the difference between the maximum and minimum data, within a data set.*

In the case of Messi and Cristiano Ronaldo, we calculate the ranges:

$$
\begin{aligned}
\mathrm{Ran}_{CR7}^{Goals} &= 69-1 = 68, & \mathrm{Ran}_{Messi}^{Goals} &= 91-3 = 88, \qquad (3.1)\\
\mathrm{Ran}_{CR7}^{Assist} &= 20-2 = 18, & \mathrm{Ran}_{Messi}^{Assist} &= 37-3 = 34, \qquad (3.2)\\
\mathrm{Ran}_{CR7}^{G+A} &= 84-3 = 81, & \mathrm{Ran}_{Messi}^{G+A} &= 113-7 = 106. \qquad (3.3)
\end{aligned}
$$

Definition 3.3. *When you have an ordered collection of data, the values that limit above at least a certain pre-established percentage of the population are called percentiles.*

We construct the next tables for the percentils:

n	Percentage	Data	Points (G+A)	straight line
10	$(0.1)\,(22) = 2.2$	2 and 3	$(2,20)$ and $(3,26)$	$y = 20 + \left[\frac{26-20}{3-2}\right](x-2)$

[7] R is a free and open-source programming language for statistical and graphical analysis. It was developed in 1993 by Robert Gentleman Ross Ihaka in the statistics department at the University of Auckland in New Zealand.

20	$(0.2)(22) = 4.4$	4 and 5	$(4, 27)$ and $(5, 35)$	$y = 27 + \left[\frac{35-27}{5-4}\right](x-4)$
30	$(0.3)(22) = 6.6$	6 and 7	$(6, 45)$ and $(7, 46)$	$y = 45 + \left[\frac{46-45}{7-6}\right](x-6)$
40	$(0.4)(22) = 8.8$	8 and 9	$(8, 47)$ and $(9, 50)$	$y = 47 + \left[\frac{50-47}{9-8}\right](x-8)$
50	$(0.5)(22) = 11$	11	$(11, 53)$	
60	$(0.6)(22) = 13.2$	13 and 14	$(12, 58)$ and $(14, 62)$	$y = 58 + \left[\frac{62-58}{14-13}\right](x-13)$
70	$(0.7)(22) = 15.4$	15 and 16	$(15, 65)$ and $(16, 66)$	$y = 65 + \left[\frac{66-65}{16-15}\right](x-15)$
80	$(0.8)(22) = 17.6$	17 and 18	$(17, 71)$ and $(18, 74)$	$y = 71 + \left[\frac{74-71}{18-17}\right](x-17)$
90	$(0.9)(22) = 19.8$	19 and 20	$(19, 76)$ and $(20, 78)$	$y = 76 + \left[\frac{78-76}{20-19}\right](x-19)$
100	$(1.0)(22) = 22$	22	$(22, 84)$	

Table 3.1: Percentils of Cristiano Ronaldo's G+A.

Evaluating we have:

$$
\begin{aligned}
(P_{10})^{G+A}_{CR7} &= 21.2, & (P_{20})^{G+A}_{CR7} &= 30.2,\\
(P_{30})^{G+A}_{CR7} &= 45.6, & (P_{40})^{G+A}_{CR7} &= 49.4,\\
(P_{50})^{G+A}_{CR7} &= 53, & (P_{60})^{G+A}_{CR7} &= 58.8,\\
(P_{70})^{G+A}_{CR7} &= 65.4, & (P_{80})^{G+A}_{CR7} &= 72.8,\\
(P_{90})^{G+A}_{CR7} &= 77.6, & (P_{100})^{G+A}_{CR7} &= 84.
\end{aligned}
$$

n	Percentage	Data	Points (G+A)	straight line
10	$(0.1)(19) = 1.9$	1 and 2	$(1, 15)$ and $(2, 40)$	$y = 20 + \left[\frac{26-20}{3-2}\right](x-1)$
20	$(0.2)(19) = 3.8$	3 and 4	$(3, 40)$ and $(4, 43)$	$y = 27 + \left[\frac{35-27}{5-4}\right](x-3)$
30	$(0.3)(19) = 5.7$	5 and 6	$(5, 46)$ and $(6, 47)$	$y = 46 + \left[\frac{47-46}{7-6}\right](x-5)$
40	$(0.4)(19) = 7.6$	7 and 8	$(7, 56)$ and $(8, 61)$	$y = 56 + \left[\frac{61-56}{8-7}\right](x-7)$
50	$(0.5)(19) = 9.5$	9 and 10	$(9, 61)$ and $(10, 65)$	$y = 61 + \left[\frac{65-61}{10-9}\right](x-9)$
60	$(0.6)(19) = 11.4$	11 and 12	$(11, 68)$ and $(12, 70)$	$y = 68 + \left[\frac{70-68}{12-11}\right](x-11)$
70	$(0.7)(19) = 13.3$	13 and 14	$(13, 77)$ and $(14, 77)$	$y = 77 + \left[\frac{77-77}{14-13}\right](x-13)$
80	$(0.8)(19) = 15.2$	15 and 16	$(15, 78)$ and $(16, 79)$	$y = 78 + \left[\frac{79-78}{16-15}\right](x-15)$
90	$(0.9)(19) = 17.1$	17 and 18	$(17, 90)$ and $(18, 96)$	$y = 90 + \left[\frac{96-90}{18-17}\right](x-17)$
100	$(1.0)(22) = 19$	19	$(19, 113)$	

Table 3.2: Percentils of Messi's G+A.

Evaluating, we have:

$$(P_{10})^{G+A}_{Messi} = 37.5, \qquad (P_{20})^{G+A}_{Messi} = 42.4, \tag{3.4}$$

$$(P_{30})^{G+A}_{Messi} = 46.7, \qquad (P_{40})^{G+A}_{Messi} = 59, \tag{3.5}$$

$$(P_{50})^{G+A}_{Messi} = 63, \qquad (P_{60})^{G+A}_{Messi} = 68.8, \tag{3.6}$$

$$(P_{70})^{G+A}_{Messi} = 77, \qquad (P_{80})^{G+A}_{Messi} = 78.2, \tag{3.7}$$

$$(P_{90})^{G+A}_{Messi} = 90.6, \qquad (P_{100})^{G+A}_{Messi} = 113. \tag{3.8}$$

n	Percentage	Data	Points (Goals)	straight line
10	$(0.1)(22) = 2.2$	2 and 3	$(2, 13)$ and $(3, 15)$	$y = 13 + \left[\frac{15-13}{3-2}\right](x-2)$
20	$(0.2)(22) = 4.4$	4 and 5	$(4, 16)$ and $(5, 25)$	$y = 16 + \left[\frac{25-16}{5-4}\right](x-4)$
30	$(0.3)(22) = 6.6$	6 and 7	$(6, 30)$ and $(7, 34)$	$y = 30 + \left[\frac{34-30}{7-6}\right](x-6)$
40	$(0.4)(22) = 8.8$	8 and 9	$(8, 35)$ and $(9, 39)$	$y = 35 + \left[\frac{39-35}{9-8}\right](x-8)$
50	$(0.5)(22) = 11$	11	$(11, 44)$	
60	$(0.6)(22) = 13.2$	13 and 14	$(13, 48)$ and $(14, 49)$	$y = 48 + \left[\frac{49-48}{14-13}\right](x-13)$
70	$(0.7)(22) = 15.4$	15 and 16	$(15, 53)$ and $(16, 54)$	$y = 65 + \left[\frac{66-65}{16-15}\right](x-15)$
80	$(0.8)(22) = 17.6$	17 and 18	$(17, 55)$ and $(18, 57)$	$y = 55 + \left[\frac{57-55}{18-17}\right](x-17)$
90	$(0.9)(22) = 19.8$	19 and 20	$(19, 60)$ and $(20, 61)$	$y = 60 + \left[\frac{61-60}{20-19}\right](x-19)$
100	$(1.0)(22) = 22$	22	$(22, 69)$	

Table 3.3: Percentils of Cristiano Ronaldo's Goals.

Evaluating we have:

$$(P_{10})^{Goals}_{CR7} = 13.4, \qquad (P_{20})^{Goals}_{CR7} = 19.6, \tag{3.9}$$
$$(P_{30})^{Goals}_{CR7} = 32.4, \qquad (P_{40})^{Goals}_{CR7} = 38.4, \tag{3.10}$$
$$(P_{50})^{Goals}_{CR7} = 44, \qquad (P_{60})^{Goals}_{CR7} = 48.2, \tag{3.11}$$
$$(P_{70})^{Goals}_{CR7} = 53.4, \qquad (P_{80})^{Goals}_{CR7} = 56.2, \tag{3.12}$$
$$(P_{90})^{Goals}_{CR7} = 60.8, \qquad (P_{100})^{Goals}_{CR7} = 69. \tag{3.13}$$

n	Percentage	Data	Points (Goals)	straight line
10	$(0.1)(19) = 1.9$	1 and 2	$(1, 12)$ and $(2, 22)$	$y = 12 + \left[\frac{22-12}{2-1}\right](x-1)$
20	$(0.2)(19) = 3.8$	3 and 4	$(3, 27)$ and $(4, 28)$	$y = 27 + \left[\frac{28-27}{4-3}\right](x-3)$
30	$(0.3)(19) = 5.7$	5 and 6	$(5, 29)$ and $(6, 31)$	$y = 29 + \left[\frac{31-29}{6-5}\right](x-5)$
40	$(0.4)(19) = 7.6$	7 and 8	$(7, 35)$ and $(8, 41)$	$y = 35 + \left[\frac{41-35}{8-7}\right](x-7)$
50	$(0.5)(19) = 9.5$	9 and 10	$(9, 43)$ and $(10, 45)$	$y = 43 + \left[\frac{45-43}{10-9}\right](x-9)$
60	$(0.6)(19) = 11.4$	11 and 12	$(11, 50)$ and $(12, 51)$	$y = 50 + \left[\frac{51-50}{12-11}\right](x-11)$
70	$(0.7)(19) = 13.3$	13 and 14	$(13, 52)$ and $(14, 54)$	$y = 52 + \left[\frac{54-52}{14-13}\right](x-13)$
80	$(0.8)(19) = 15.2$	15 and 16	$(15, 58)$ and $(16, 59)$	$y = 58 + \left[\frac{59-58}{16-15}\right](x-15)$
90	$(0.9)(19) = 17.1$	17 and 18	$(17, 59)$ and $(18, 60)$	$y = 59 + \left[\frac{60-59}{18-17}\right](x-17)$
100	$(1.0)(19) = 19$	19	$(19, 91)$	

Table 3.4: Percentils of Messi's Goals.

$$(P_{10})^{Goals}_{Messi} = 21, \qquad (P_{20})^{Goals}_{Messi} = 27.8, \tag{3.14}$$
$$(P_{30})^{Goals}_{Messi} = 30.4, \qquad (P_{40})^{Goals}_{Messi} = 38.6, \tag{3.15}$$
$$(P_{50})^{Goals}_{Messi} = 44, \qquad (P_{60})^{Goals}_{Messi} = 50.4, \tag{3.16}$$
$$(P_{70})^{Goals}_{Messi} = 52.6. \qquad (P_{80})^{Goals}_{Messi} = 58.2, \tag{3.17}$$
$$(P_{90})^{Goals}_{Messi} = 59.1, \qquad (P_{100})^{Goals}_{Messi} = 91. \tag{3.18}$$

n	Percentage	Data	Points (Assists)	straight line
10	$(0.1)(22) = 2.2$	2 and 3	$(2, 4)$ and $(3, 4)$	$y = 4 + \left[\frac{4-4}{3-2}\right](x-2)$
20	$(0.2)(22) = 4.4$	4 and 5	$(4, 7)$ and $(5, 7)$	$y = 7 + \left[\frac{7-7}{5-4}\right](x-4)$
30	$(0.3)(22) = 6.6$	6 and 7	$(6, 7)$ and $(7, 10)$	$y = 7 + \left[\frac{10-7}{7-6}\right](x-6)$
40	$(0.4)(22) = 8.8$	8 and 9	$(8, 11)$ and $(9, 11)$	$y = 11 + \left[\frac{11-11}{9-8}\right](x-8)$
50	$(0.5)(22) = 11$	11	$(11, 12)$	
60	$(0.6)(22) = 13.2$	13 and 14	$(13, 13)$ and $(14, 14)$	$y = 13 + \left[\frac{14-13}{14-13}\right](x-13)$
70	$(0.7)(22) = 15.4$	15 and 16	$(15, 14)$ and $(16, 15)$	$y = 14 + \left[\frac{15-14}{16-15}\right](x-15)$
80	$(0.8)(22) = 17.6$	17 and 18	$(17, 16)$ and $(18, 16)$	$y = 16 + \left[\frac{16-16}{18-17}\right](x-17)$
90	$(0.9)(22) = 19.8$	19 and 20	$(19, 17)$ and $(20, 18)$	$y = 17 + \left[\frac{18-17}{20-19}\right](x-19)$
100	$(1.0)(22) = 22$	22	$(22, 20)$	

Table 3.5: Percentils of Cristiano Ronaldo's Assists.

Evaluating we have:

$$(P_{10})^{Assist}_{CR7} = 4, \qquad (P_{20})^{Assist}_{CR7} = 7,$$
$$(P_{30})^{Assist}_{CR7} = 8.8, \qquad (P_{40})^{Assist}_{CR7} = 11,$$
$$(P_{50})^{Assist}_{CR7} = 12, \qquad (P_{60})^{Assist}_{CR7} = 13.2,$$
$$(P_{70})^{Assist}_{CR7} = 14.4, \qquad (P_{80})^{Assist}_{CR7} = 16,$$
$$(P_{90})^{Assist}_{CR7} = 17.8, \qquad (P_{100})^{Assist}_{CR7} = 20.$$

n	Percentage	Data	Points (Assists)	straight line
10	$(0.1)(19) = 1.9$	1 and 2	$(1, 4)$ and $(2, 12)$	$y = 4 + \left[\frac{12-4}{2-1}\right](x-1)$
20	$(0.2)(19) = 3.8$	3 and 4	$(3, 12)$ and $(4, 15)$	$y = 12 + \left[\frac{15-12}{4-3}\right](x-3)$
30	$(0.3)(19) = 5.7$	5 and 6	$(5, 16)$ and $(6, 16)$	$y = 16 + \left[\frac{16-16}{6-5}\right](x-5)$
40	$(0.4)(19) = 7.6$	7 and 8	$(7, 17)$ and $(8, 18)$	$y = 17 + \left[\frac{18-17}{8-7}\right](x-7)$
50	$(0.5)(19) = 9.5$	9 and 10	$(9, 18)$ and $(10, 18)$	$y = 18 + \left[\frac{18-18}{10-9}\right](x-9)$
60	$(0.6)(19) = 11.4$	11 and 12	$(11, 18)$ and $(12, 19)$	$y = 18 + \left[\frac{19-18}{12-11}\right](x-11)$
70	$(0.7)(19) = 13.3$	13 and 14	$(13, 21)$ and $(14, 22)$	$y = 21 + \left[\frac{22-21}{14-13}\right](x-13)$
80	$(0.8)(19) = 15.2$	15 and 16	$(15, 26)$ and $(16, 26)$	$y = 26 + \left[\frac{26-26}{16-15}\right](x-15)$

90	$(0.9)(19) = 17.1$	17 and 18	$(17, 30)$ and $(18, 31)$	$y = 30 + \left[\frac{31-30}{18-17}\right](x - 17)$
100	$(1.0)(19) = 19$	19	$(19, 37)$	

Table 3.6: Percentils of Messi's Assists.

Evaluating, we have:

$$\begin{aligned} (P_{10})^{Assist}_{Messi} &= 11.2, & (P_{20})^{Assist}_{Messi} &= 14.4, \\ (P_{30})^{Assist}_{Messi} &= 16, & (P_{40})^{Assist}_{Messi} &= 17.6, \\ (P_{50})^{Assist}_{Messi} &= 18, & (P_{60})^{Assist}_{Messi} &= 18.4, \\ (P_{70})^{Assist}_{Messi} &= 21.3, & (P_{80})^{Assist}_{Messi} &= 26, \\ (P_{90})^{Assist}_{Messi} &= 30.1, & (P_{100})^{Assist}_{Messi} &= 37. \end{aligned}$$

Definition 3.4. *Quartiles*[8] *are the three values that divide the ordered data set into four parts.*

3.1. CR7's goals. We ordered the data from smallest to largest

$$\{Goals\}_{CR7} = \{1, 5, 13, 15, 16, 25, 30, 34, 35, 39, 43, 44, 47, 48, 49, 53, 54, 55, 57, 60, 61, 63, 69\}.$$

Since there are 23 data points, Q_2 is the mediane

$$Q_2 = \text{Med}^{Goals}_{CR7} = 44, \tag{3.19}$$

and it occupies the 12th place. To calculate the other quartiles we must consider two groups.

First group. Data that occupies positions less than or equal to 12. That is, from the first to the eleventh:

$$\{Goals\}_{CR7} = \{1, 5, 13, 15, 16, 25, 30, 34, 35, 39, 43\}.$$

The first quartile is the median of these data:

$$Q_1 = 25. \tag{3.20}$$

Second group. Data that occupy positions greater than or equal to 12; that is, from the thirteenth to the last

$$\{Goals\}_{CR7} = \{47, 48, 49, 53, 54, 55, 57, 60, 61, 63, 69\}.$$

The third quartile is the median of these data

$$Q_3 = 55. \tag{3.21}$$

3.2. Messi's goals. We ordered the data from smallest to largest

$$\{Goals\}_{Messi} = \{3, 12, 22, 27, 28, 29, 31, 35, 41, 43, 45, 50, 51, 52, 54, 58, 59, 59, 60, 91\},$$

Since there are 20 data points, Q_2 is the mediane

$$Q_2 = \text{Med}^{Goals}_{Messi} = \frac{43 + 45}{2} = 44, \tag{3.22}$$

and it occupies the place

$$q_2 = \frac{10 + 11}{2} = 10.5$$

To calculate the other quartiles we must consider two groups.

[8] Maurice Kendall (1907-1983), an English mathematician, first used the term quartile in 1940.

First group. Data that occupies positions less than or equal to 10. That is, from the first to the tenth:

$$\{Goals\}_{Messi} = \{3, 12, 22, 27, 28, 29, 31, 35, 41, 43\}\,.$$

The first quartile is the median of these data:

$$Q_1 = \frac{28 + 29}{2} = 28.5. \tag{3.23}$$

Second group. Data that occupy positions greater than or equal to 11; that is, from the eleventh to the last

$$\{Goals\}_{Messi} = \{45, 50, 51, 52, 54, 58, 59, 59, 60, 91\}\,.$$

The third quartile is the median of these data

$$Q_3 = \frac{54 + 58}{2} = 56. \tag{3.24}$$

3.3. CR7's assists. We ordered the data from smallest to largest

$$\{Assists\}_{CR7} = \{2, 4, 4, 4, 4, 7, 7, 10, 11, 11, 12, 12, 13, 13, 14, 15, 16, 16, 17, 18, 18, 20\}\,.$$

Since there are 23 data points, Q_2 is the mediane

$$Q_2 = \mathrm{Med}_{CR7}^{Assist} = 12, \tag{3.25}$$

and it occupies the 12th place. To calculate the other quartiles we must consider two groups.

First group. Data that occupies positions less than or equal to 12. That is, from the first to the eleventh:

$$\{Assists\}_{CR7} = \{2, 4, 4, 4, 4, 7, 7, 10, 11, 11, 12, 12\}\,,$$

The first quartile is the median of these data:

$$Q_1 = 7. \tag{3.26}$$

Second group. Data that occupy positions greater than or equal to 12; that is, from the thirteenth to the last

$$\{Assists\}_{CR7} = \{12, 13, 13, 14, 15, 16, 16, 17, 18, 18, 20\}\,,$$

The third quartile is the median of these data

$$Q_3 = 16. \tag{3.27}$$

3.4. Messi's assists. We ordered the data from smallest to largest

$$\{Assists\}_{Messi} = \{3, 4, 12, 12, 15, 16, 16, 17, 18, 18, 18, 18, 19, 21, 22, 26, 30, 31, 37\}\,.$$

Since there are 20 data points, Q_2 is the mediane

$$Q_2 = \mathrm{Med}_{Messi}^{Assist} = 18, \tag{3.28}$$

and it occupies the place

$$q_2 = \frac{10 + 11}{2} = 10.5$$

To calculate the other quartiles we must consider two groups.

First group. Data that occupies positions less than or equal to 10. That is, from the first to the tenth:

$$\{Assists\}_{Messi} = \{3, 4, 12, 12, 15, 16, 16, 17, 18, 18\}\,.$$

The first quartile is the median of these data:

$$Q_1 = \frac{15 + 16}{2} = 15.5. \tag{3.29}$$

Second group. Data that occupy positions greater than or equal to 11; that is, from the eleventh to the last

$$\{Assists\}_{Messi} = \{18, 18, 18, 19, 21, 22, 26, 30, 31, 37\}.$$

The third quartile is the median of these data

$$Q_3 = \frac{21+22}{2} = 21.5. \tag{3.30}$$

3.5. CR7's goals+assists. We ordered the data from smallest to largest

$$\{G+A\}_{CR7} = \{3, 9, 20, 26, 27, 35, 45, 46, 47, 50, 51, 53, 54, 58, 62, 65, 66, 71, 74, 76, 78, 81, 84\}.$$

Since there are 23 data points, Q_2 is the mediane

$$Q_2 = \text{Med}_{CR7}^{G+A} = 53, \tag{3.31}$$

$$\text{Med}_{Messi}^{G+A} = \frac{61+65}{2} = 63. \tag{3.32}$$

and it occupies the 12th place. To calculate the other quartiles we must consider two groups.

First group. Data that occupies positions less than or equal to 12. That is, from the first to the eleventh:

$$\{G+A\}_{CR7} = \{3, 9, 20, 26, 27, 35, 45, 46, 47, 50, 51\}.$$

The first quartile is the median of these data:

$$Q_1 = 35. \tag{3.33}$$

Second group. Data that occupy positions greater than or equal to 12; that is, from the thirteenth to the last

$$\{G+A\}_{CR7} = \{54, 58, 62, 65, 66, 71, 74, 76, 78, 81, 84\}.$$

The third quartile is the median of these data

$$Q_3 = 71. \tag{3.34}$$

3.6. Messi's goals+assists. We ordered the data from smallest to largest

$$\{G+A\}_{Messi} = \{7, 15, 40, 40, 43, 46, 47, 56, 61, 61, 65, 68, 70, 77, 77, 78, 79, 90, 96, 113\}.$$

Since there are 20 data points, Q_2 is the mediane

$$Q_2 = \text{Med}_{Messi}^{G+A} = \frac{61+65}{2} = 63. \tag{3.35}$$

and it occupies the place

$$q_2 = \frac{10+11}{2} = 10.5$$

To calculate the other quartiles we must consider two groups.

First group. Data that occupies positions less than or equal to 10. That is, from the first to the tenth:

$$\{G+A\}_{Messi} = \{7, 15, 40, 40, 43, 46, 47, 56, 61, 61\}.$$

The first quartile is the median of these data:

$$Q_1 = \frac{43+46}{2} = 44.5. \tag{3.36}$$

Second group. Data that occupy positions greater than or equal to 11; that is, from the eleventh to the last

$$\{G + A\}_{Messi} = \{65, 68, 70, 77, 77, 78, 79, 90, 96, 113\}.$$

The third quartile is the median of these data

$$Q_3 = \frac{77 + 78}{2} = 77.5. \tag{3.37}$$

Definition 3.5. *The interquartile range (RIC) is a measure of dispersion that is not affected by aberrant data, and to determine it, it is necessary to consider the size of the range of data located between the first and third quartiles, that is,*

$$RIC = Q_3 - Q_1. \tag{3.38}$$

The RICs are:

$$\begin{aligned} RIC^{Goals}_{CR7} &= 55 - 25 = 30, & RIC^{Goals}_{Messi} &= 56 - 28.5 = 28.5, \\ RIC^{Assist}_{CR7} &= 16 - 7 = 9, & RIC^{Assist}_{Messi} &= 21.5 - 15.5 = 6, \\ RIC^{G+A}_{CR7} &= 71 - 35 = 36, & RIC^{G+A}_{Messi} &= 77.5 - 44.5 = 33. \end{aligned}$$

Definition 3.6. *The variance*[9]*, V, of a data set is the average of the squares of its deviations from the mean.* $\mu = \langle x \rangle$*, that is:*

$$V = \frac{1}{N} \sum_{k=1}^{N} (x_k - \mu)^2. \tag{3.39}$$

The variances of CR7 and Messi are:

$$V^{Goals}_{CR7} = \frac{1}{23} \sum_{k=1}^{23} (x_k - \mu)^2 = \frac{8371.32}{23} = 363.97, \tag{3.40}$$

$$V^{Goals}_{Messi} = \frac{1}{20} \sum_{k=1}^{20} (x_k - \mu)^2 = \frac{7535}{20} = 376.75, \tag{3.41}$$

$$V^{Assist}_{CR7} = \frac{1}{23} \sum_{k=1}^{23} (x_k - \mu)^2 = \frac{579.739}{23} = 25.20, \tag{3.42}$$

$$V^{Assist}_{Messi} = \frac{1}{20} \sum_{k=1}^{20} (x_k - \mu)^2 = \frac{1318.445}{23} = 65.92, \tag{3.43}$$

$$V^{G+A}_{CR7} = \frac{1}{23} \sum_{k=1}^{23} (x_k - \mu)^2 = \frac{13117.175}{23} = 572.92, \tag{3.44}$$

$$V^{G+A}_{Messi} = \frac{1}{20} \sum_{k=1}^{20} (x_k - \mu)^2 = \frac{14010.568}{23} = 700.52. \tag{3.45}$$

The standard deviation are:

$$\sigma^{Goals}_{CR7} = \sqrt{\frac{1}{23} \sum_{k=1}^{23} (x_k - \mu)^2} = \sqrt{363.97} = 19.07, \tag{3.46}$$

[9]The English mathematician Ronald Fisher (1890-1962) first used the term variance in the article entitled “The correlation between relatives on the supposition of mendelian inheritance”, published in 1919.

$$\sigma_{Messi}^{Goals} = \sqrt{\frac{1}{20}\sum_{k=1}^{20}(x_k-\mu)^2} = \sqrt{376.75} = 19.41, \tag{3.47}$$

$$\sigma_{CR7}^{Assist} = \sqrt{\frac{1}{23}\sum_{k=1}^{23}(x_k-\mu)^2} = \sqrt{25.20} = 5.01, \tag{3.48}$$

$$\sigma_{Messi}^{Assist} = \sqrt{\frac{1}{20}\sum_{k=1}^{20}(x_k-\mu)^2} = \sqrt{65.92} = 8.12, \tag{3.49}$$

$$\sigma_{CR7}^{G+A} = \sqrt{\frac{1}{23}\sum_{k=1}^{23}(x_k-\mu)^2} = \sqrt{572.92} = 23.93, \tag{3.50}$$

$$\sigma_{Messi}^{G+A} = \sqrt{\frac{1}{20}\sum_{k=1}^{20}(x_k-\mu)^2} = \sqrt{700.52} = 26.46. \tag{3.51}$$

TChebysheff's[10] theorem. . Let X be a random variable with finite mean μ and variance σ^2. Then, for any $k > 0$,

$$P(|X-\mu| < k\sigma) \geq 1 - \frac{1}{k^2} \tag{3.52}$$

y

$$P(|X-\mu| \geq k\sigma) < \frac{1}{k^2}. \tag{3.53}$$

Proof: We will give the proof for a continuous random variable. The proof for the discrete case proceeds in a similar way. Let $f(x)$ denote the density function of X. Then,

$$V(X) = \sigma^2 = \int_{-\infty}^{+\infty}(x-\mu)^2 f(x)\,dx = \left(\int_{-\infty}^{\mu-k\sigma} + \int_{\mu-k\sigma}^{\mu+k\sigma} + \int_{\mu+k\sigma}^{+\infty}\right)(x-\mu)^2 f(x)\,dx$$

or

$$V(X) = \int_{-\infty}^{\mu-k\sigma}(x-\mu)^2 f(x)\,dx + \int_{\mu-k\sigma}^{\mu+k\sigma}(x-\mu)^2 f(x)\,dx + \int_{\mu+k\sigma}^{+\infty}(x-\mu)^2 f(x)\,dx.$$

The second integral is always greater than or equal to zero and $(x-\mu)^2 \geq k^2\sigma^2$ for all values of x between the limits of integration for the first and third integrals; that is, the integration regions are in the tails of the density function and include only values of x for which $(x-\mu)^2 \geq k^2\sigma^2$. Substitute the second integral with zero and substitute $k^2\sigma^2$ for $(x-\mu)^2$ in the first and third integrals to obtain the inequality

$$\sigma^2 \geq \int_{-\infty}^{\mu-k\sigma} k^2\sigma^2 f(x)\,dx + \int_{\mu+k\sigma}^{+\infty} k^2\sigma^2 f(x)\,dx$$

or

$$\sigma^2 \geq k^2\sigma^2 P(X \leq \mu - k\sigma) + k^2\sigma^2 P(X \geq \mu + k\sigma)$$

[10]The Russian Psfnuty Lvovich Tchebysheff is one of the most celebrated mathematicians of the 19th century. His research focused primarily on mechanisms and the theory of function approximation, number theory, probability theory, and integration theory.

Dividing by $k^2\sigma^2$, we get

$$P\left(|X-\mu|\geq k\sigma\right)\leq \frac{1}{k^2}$$

o bien, lo que es equivalente,

$$P\left(|X-\mu|< k\sigma\right)\geq 1-\frac{1}{k^2}.$$

□

If we take $k=2$ in Tchebysheff's theorem, we obtain:

$$\left(\langle x\rangle_{CR7}^{Goals}-2\sigma_{CR7}^{Goals},\langle x\rangle_{CR7}^{Goals}+2\sigma_{CR7}^{Goals}\right)=\left(39.82-2\left(19.07\right),39.82+2\left(19.07\right)\right)=\left(1.68,77.96\right),$$

$$\Rightarrow 1\notin\left(1.68,77.96\right).$$

$$\left(\langle x\rangle_{Messi}^{Goals}-2\sigma_{Messi}^{Goals},\langle x\rangle_{Messi}^{Goals}+2\sigma_{Messi}^{Goals}\right)=\left(42.5-2\left(19.41\right),42.5+2\left(19.41\right)\right)=\left(3.68,81.32\right),$$

$$3\notin\left(3.68,81.32\right),$$

$$91\notin\left(3.68,81.32\right).$$

$$\left(\langle x\rangle_{CR7}^{Assists}-2\sigma_{CR7}^{Assists},\langle x\rangle_{CR7}^{Assists}+2\sigma_{CR7}^{Assists}\right)=\left(11.52-2\left(5.01\right),11.52+2\left(5.01\right)\right)=\left(1.5,21.54\right)$$

$$\Rightarrow \forall x\in\left(1.5,21.54\right).$$

$$\left(\langle x\rangle_{Messi}^{Assists}-2\sigma_{Messi}^{Assists},\langle x\rangle_{Messi}^{Assists}+2\sigma_{Messi}^{Assists}\right)=\left(18.95-2\left(8.12\right),18.95+2\left(8.12\right)\right)=\left(2.71,35.19\right),$$

$$\Rightarrow 37\notin\left(2.71,35.19\right).$$

$$\left(\langle x\rangle_{CR7}^{G+A}-2\sigma_{CR7}^{G+A},\langle x\rangle_{CR7}^{G+A}+2\sigma_{CR7}^{G+A}\right)=\left(51.34-2\left(23.93\right),51.34+2\left(23.93\right)\right)=\left(3.48,99.2\right),$$

$$\Rightarrow 3\notin\left(3.48,99.2\right).$$

$$\left(\langle x\rangle_{Messi}^{G+A}-2\sigma_{Messi}^{G+A},\langle x\rangle_{Messi}^{G+A}+2\sigma_{Messi}^{G+A}\right)=\left(61.45-2\left(26.46\right),61.45+2\left(26.46\right)\right)=\left(8.43,114.37\right),$$

$$\Rightarrow 7\notin\left(8.43,114.37\right).$$

Note that all data, except those marked outside the intervals, are less than 2 standard deviations away from the mean.

4. Concluding remarks. We put into practice the methods, tools, and resources for classifying statistical data and graphically representing it. We analyzed the datasets of Cristiano Ronaldo and Lionel Messi with respect to their goals, assists, and goal contributions. We also reviewed the importance of measures of central tendency, such as the arithmetic mean, harmonic mean, geometric mean, median, mode, and midrange. Furthermore, we studied the variation present in the data associated with a statistical phenomenon, the quantification of which is carried out using measures of dispersion, such as range, percentiles, variance, and standard deviation, in the problem presented here. The most commonly used measure of dispersion is the standard deviation.It should be noted that in all statistical measures or parameters Messi is superior to Cristiano Ronaldo, therefore, Messi is a more decisive football player than CR7.

It tells us how centered the data is, that is, how representative the mean is of the total data. If the standard deviation is small, most of the data is close to the average; however, if it is large, the data is more evenly distributed. Finally, we studied Chebysheff's theorem, which allows us, given a number $k>1$, to determine the percentage of data that are less than k times the standard deviation of the mean.

A real value of Tchebysheff's theorem is that it makes it possible to find bounds on probabilities that would ordinarily have to be obtained through tedious mathematical manipulations (integration or summation). Furthermore, we can often obtain means and variances of random variables without specifying the distribution of the variable. In situations like this, Tchebysheff's theorem still provides meaningful bounds on probabilities of interest.

Author Contributions. The formal analysis, research, and writing – original draft preparation was completed by Mississippi Valenzuela. Computations were made by Ingrid D. Castro, Wendy G. Hernández, Isabela Solis, María C. Guerra, Linda J. Murillo, Wendy N. Del Rio, Beatriz A. Rios, Maximiliano Martínez. All authors have read and agreed to the published version of the manuscript

Acknowledgment. MV would like to thank L.C.F.D. Adan Dionicio Muñoz Montes for the useful discussions and comments onthe topics.

Conflicts of interest. The authors declares no conflict of interest.

Abbreviations. The following abbreviations are used in this manuscript:

SelMat	Selecciones Matematicas
DOAJ	Directory of open access journals
TLA	Three letter acronym
LD	Linear dichroism

ORCID and License

Mississippi Valenzuela https://orcid.org/0000-0002-8532-229X
Ingrid D. Castro https://orcid.org/0009-0003-6792-3782
Wendy G. Hernández https://orcid.org/0009-0005-5350-9470
Isabela Solis https://orcid.org/0009-0001-3070-9315
María C. Guerra https://orcid.org/0009-0003-0172-1934
Wendy N. Del Rio https://orcid.org/0009-0005-1582-738X
Beatriz A. Rios https://orcid.org/0009-0004-7311-5084
Maximiliano Martínez https://orcid.org/0000-0001-8033-7572

References

[1] Devore JL. Probabilidad y estadística para ingenieria y ciencias. Seventh edition. Cengage Learning. 2008.
[2] Durrett R. Elementary Probability for Applications. Cambrige. 2009.
[3] Häggströng O. Finite Markov Chains and Algorithmic Applications. Cambrige University Press, 2003.
[4] Isaac R, The pleasures of Probability. Undergraduate Texts in Mathematics. Springer. USA. 1995.
[5] James F. Statistical Methods in Experimental Physics. Second edition. World Scientific. Singapur. 2006.
[6] Kolmogorov AN, Grundbegriffe der Wahrscheinlichkeitsrechnung. Springer. 1933; Fundations of the Theory of Probability. Second edition. Chelsea. 1956.
[7] Maibaum G. Teoría de probabilidades y estadística matemática. Pueblo y educación. 1976.
[8] Valenzuela M. Pruebas Estadísticas en la física del bosón de Higgs. Tesis de Maestría en Matemáticas Aplicadas, Universidad Autónoma de Zacatecas, México, february 2018.
[9] Valenzuela M. Pruebas estadísticas en física de partículas. Revista de Investigación de Física. 2024; 27(3): 50-65. Doi: https://doi.org/10.15381/rif.v27i3.23824
[10] Valenzuela M. Corrección: Pruebas estadísticas en física de partículas. Revista de Investigación de Física. 2025; 28(2): 51-55. Doi: https://doi.org/10.15381/rif.v28i2.30414
[11] Ihaka R and Gentleman R. R: A Language for Data Analysis and Graphics. Journal of Computational and Graphical Statistics. 1996; 5(3): 299-314.
[12] Valenzuela M. Fourier analysis with R. Selecciones Matemáticas. 2024; 11(01): 189-210. Doi:https://doi.org/10.17268/sel.mat.2024.01.11