Escola Tècnica Superior d'Enginyeria Informàtica
Universitat Politècnica de València

THE POWER OF MUSIC

Proyecto final de primer año

Grado en Ciencia de Datos

Autores: Diego Cotino Bolufer
Miguel Camuñas Castelló
Marcos Carrasco Panadero
Pablo Gil Martínez
Sergio Samaniego Hernández

Tutor: Jose Manuel Gil

1ºB1

18/06/2024

Annexes

Annex B (Unidimensional analysis)

Tools Used for the Qualitative Variables Analysis

In addition to defining the type of qualitative variable, for its analysis the frequency table has been used to determine the number of individuals that have each value of the variable, its absolute frequency; and how much that quantity represents in the total of the sample, its relative frequency.

Moreover, for a more visual representation of the result, the bar chart has been used to show the absolute frequency and the pie chart to show the relative frequency.

Tools Used for the Quantitative Variables Analysis

In addition to defining the type of quantitative variable, for its analysis the frequency table has been used to determine the number of individuals that have each value of the variable, its absolute frequency; and how much that quantity represents in the total of the sample, its relative frequency. In order to do this table, the quantitative variables have been discretized, if it had been necessary.

First of all, the Box and Whisker chart has been used to find out if there is atypical data in the sample. Furthermore, with the objective of defining the form of the values of the variable, the standardized asymmetry coefficient of Fisher has been used to determine the asymmetry of the variable. If the quantitative variable was continuous, in case of being symmetric and having a normal distribution too, the standardized kurtosis coefficient has been used to determine its aiming or fussiness.

Regarding the position of the values of the variable, the parameters used have been the median, the quartiles and, if the sample was symmetric, the winsorized mean too. In regard to the dispersion of the values, the interquartile range has been used and, being the sample symmetric, also the winsorized standard deviation. The parameters mentioned before are robust, extreme values in the sample do not affect them.

Besides, the bar chart and the scatter plot, being the quantitative variables discrete, and also the histogram and the normal probability graph, being continuous, have been used to represent visually the form, position and dispersion of the values of the sample.

Anxiety

This variable indicates the self-perception of the level of anxiety. It is a discrete quantitative variable, whose values range from 0 to 10.

The frequency table of the variable is the next:

Tabla de Frecuencia para Anxiety

Clase	Valor	Frecuencia	Frecuencia Relativa	Frecuencia Acumulada	Frecuencia Rel. acum.
1	0	26	0,0444	26	0,0444
2	1	21	0,0358	47	0,0802
3	2	34	0,0580	81	0,1382
4	3	55	0,0939	136	0,2321
5	4	47	0,0802	183	0,3123
6	5	42	0,0717	225	0,3840
7	6	73	0,1246	298	0,5085
8	7	96	0,1638	394	0,6724
9	8	92	0,1570	486	0,8294
10	9	43	0,0734	529	0,9027
11	10	57	0,0973	586	1,0000

Figure 1

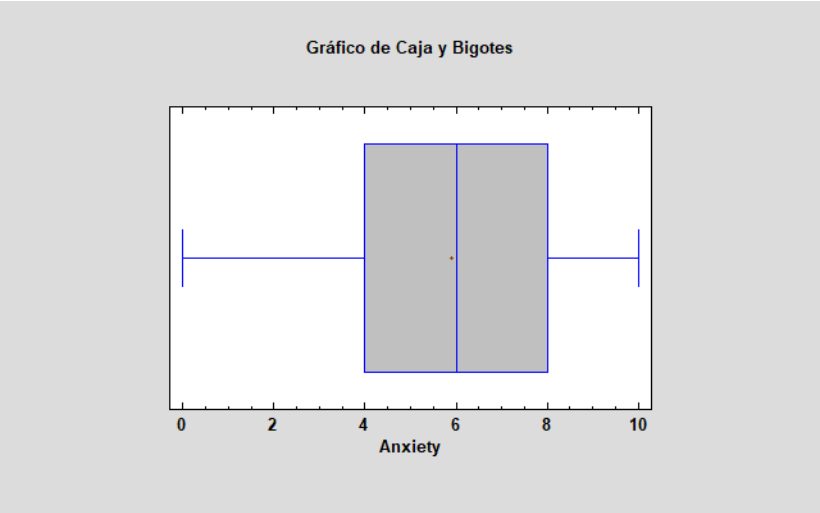

Figure 2

As it can be seen in the Box and Whisker chart there are not atypical data in the sample. Furthermore, to describe its form, the standardized asymmetry coefficient of Fisher of –4,21 indicates a negative asymmetry in the variable. It can be seen better in the bar chart and the scatter plot.

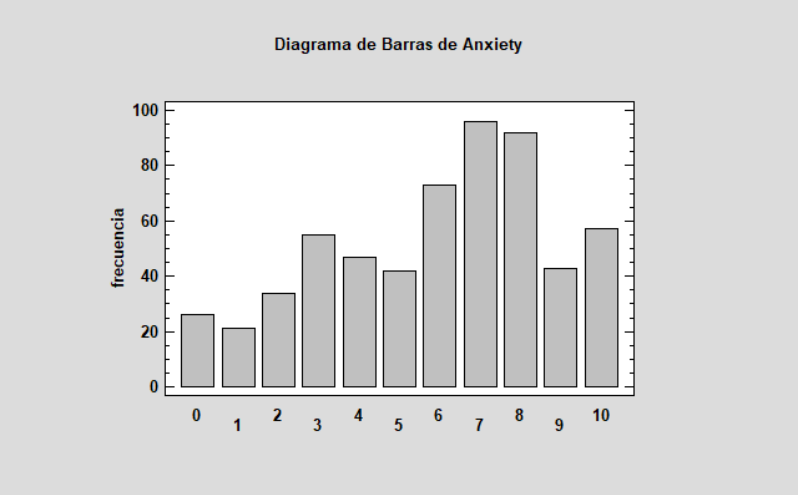

Figure 3

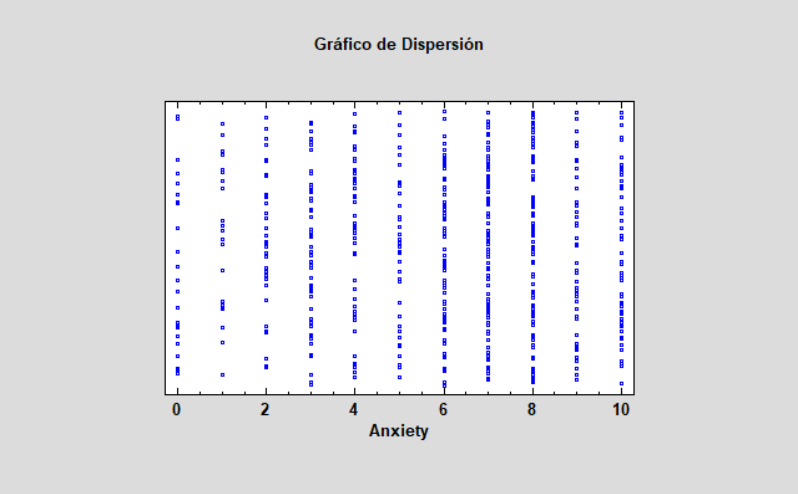

Figure 4

Regarding the position of the values, the median is 6, the lower quartile is 4 and the upper quartile is 8. Besides, about the dispersion of the sample, the interquartile range is 4.

Depression

This variable indicates the self-perception of the level of depression. It is a discrete quantitative variable, whose values range from 0 to 10.

The frequency table of the variable is the next:

Tabla de Frecuencia para Depression

Clase	Valor	Frecuencia	Frecuencia Relativa	Frecuencia Acumulada	Frecuencia Rel. acum.
1	0	65	0,1109	65	0,1109
2	1	30	0,0512	95	0,1621
3	2	71	0,1212	166	0,2833
4	3	44	0,0751	210	0,3584
5	4	53	0,0904	263	0,4488
6	5	43	0,0734	306	0,5222
7	6	76	0,1297	382	0,6519
8	7	75	0,1280	457	0,7799
9	8	63	0,1075	520	0,8874
10	9	32	0,0546	552	0,9420
11	10	34	0,0580	586	1,0000

Figure 5

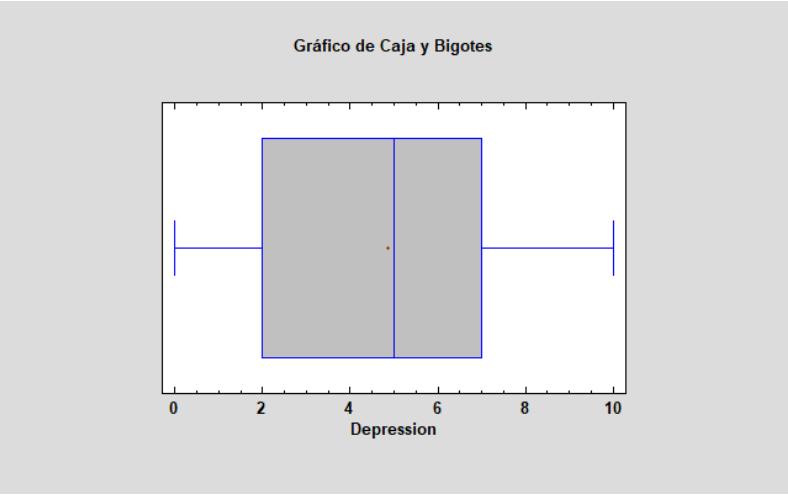

Figure 6

As it can be seen in the Box and Whisker chart there are not atypical data in the sample. Furthermore, to describe its form, the standardized asymmetry coefficient of Fisher of –0,86 indicates a symmetry in the variable. It can be seen better in the bar chart and the scatter plot.

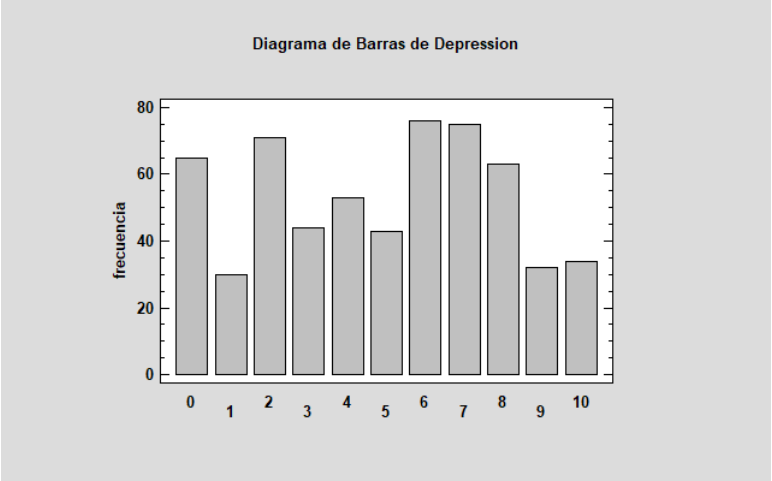

Figure 7

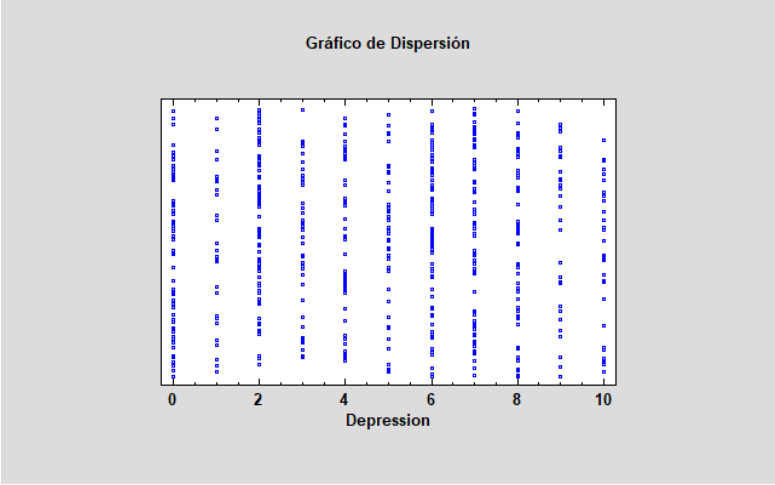

Figure 8

Regarding the position of the values, the median is 5, the lower quartile is 2 and the upper quartile is 7. Besides, about the dispersion of the sample, the interquartile range is 5.

Insomnia

This variable indicates the self-perception of the level of insomnia. It is a discrete quantitative variable, whose values range from 0 to 10.

The frequency table of the variable is the next:

Tabla de Frecuencia para Insomnia

Clase	Valor	Frecuencia	Frecuencia Relativa	Frecuencia Acumulada	Frecuencia Rel. acum.
1	0	121	0,2065	121	0,2065
2	1	62	0,1058	183	0,3123
3	2	64	0,1092	247	0,4215
4	3	58	0,0990	305	0,5205
5	4	41	0,0700	346	0,5904
6	5	47	0,0802	393	0,6706
7	6	52	0,0887	445	0,7594
8	7	52	0,0887	497	0,8481
9	8	43	0,0734	540	0,9215
10	9	22	0,0375	562	0,9590
11	10	24	0,0410	586	1,0000

Figure 9

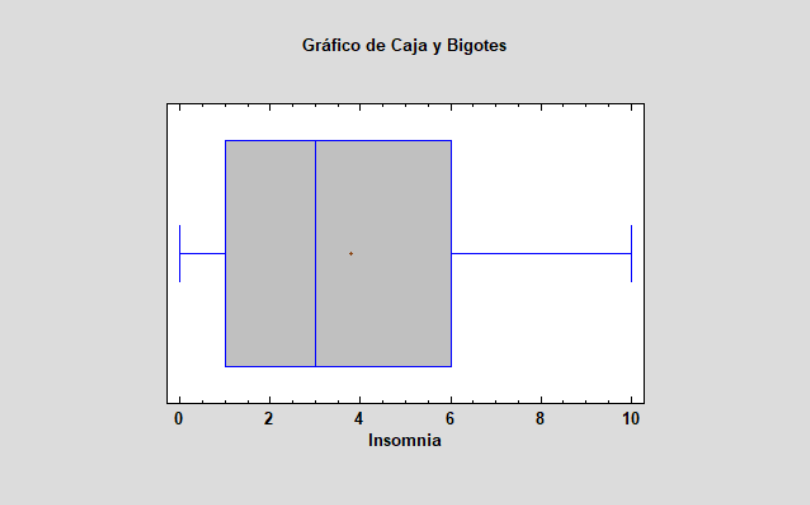

Figure 10

As it can be seen in the Box and Whisker chart there are not atypical data in the sample. Furthermore, to describe its form, the standardized asymmetry coefficient of Fisher of 3,48 indicates a positive asymmetry in the variable. It can be seen better in the bar chart and the scatter plot.

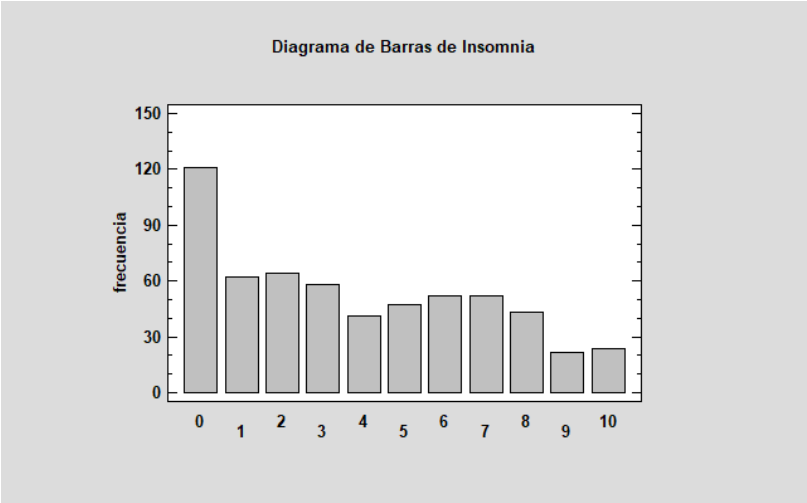

Figure 11

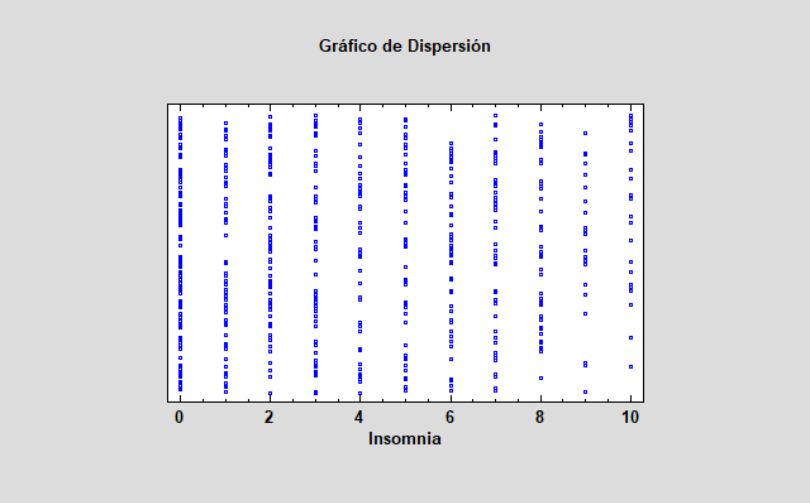

Figure 12

Regarding the position of the values, the median is 3, the lower quartile is 1 and the upper quartile is 6. Besides, about the dispersion of the sample, the interquartile range is 5.

OCD

This variable indicates the self-perception of the level of obsessive-compulsive disorder. It is a discrete quantitative variable, whose values range from 0 to 10.

The frequency table of the variable is the next:

Tabla de Frecuencia para OCD

Clase	Valor	Frecuencia	Frecuencia Relativa	Frecuencia Acumulada	Frecuencia Rel. acum.
1	0	200	0,3413	200	0,3413
2	1	78	0,1331	278	0,4744
3	2	76	0,1297	354	0,6041
4	3	50	0,0853	404	0,6894
5	4	30	0,0512	434	0,7406
6	5	45	0,0768	479	0,8174
7	6	27	0,0461	506	0,8635
8	7	27	0,0461	533	0,9096
9	8	26	0,0444	559	0,9539
10	9	11	0,0188	570	0,9727
11	10	16	0,0273	586	1,0000

Figure 13

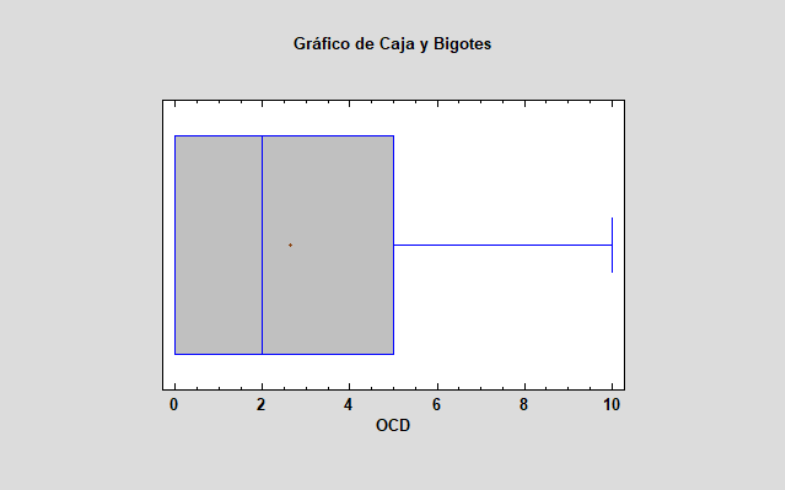

Figure 14

As it can be seen in the Box and Whisker chart there are not atypical data in the sample. Furthermore, to describe its form, the standardized asymmetry coefficient of Fisher of 9,45 indicates a great positive asymmetry in the variable. It can be seen better in the bar chart and the scatter plot.

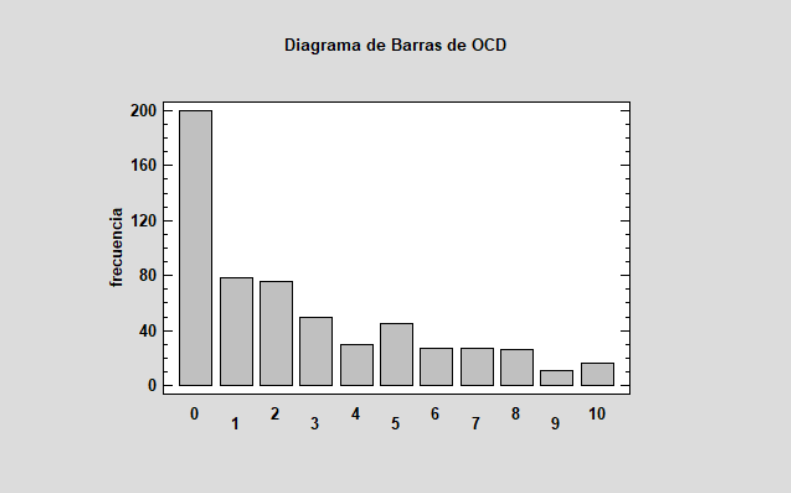

Figure 15

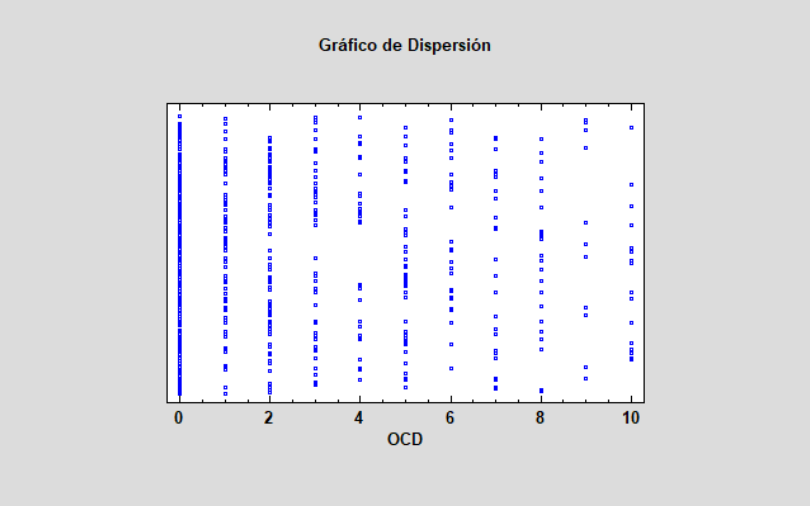

Figure 15

Regarding the position of the values, the median is 2, the lower quartile is 0 and the upper quartile is 5. Besides, about the dispersion of the sample, the interquartile range is 5.

Hours per day

This variable indicates the number of hours that each individual listens to music. It is a discrete quantitative variable, and its values go from 0 to 10.

The frequency table of the variable is the next:

Tabla de Frecuencia para Hours per day

Clase	*Valor*	*Frecuencia*	*Frecuencia Relativa*	*Frecuencia Acumulada*	*Frecuencia Rel. acum.*
1	0	3	0,0051	3	0,0051
2	1	99	0,1689	102	0,1741
3	2	133	0,2270	235	0,4010
4	3	108	0,1843	343	0,5853
5	4	76	0,1297	419	0,7150
6	5	63	0,1075	482	0,8225
7	6	41	0,0700	523	0,8925
8	7	14	0,0239	537	0,9164
9	8	27	0,0461	564	0,9625
10	9	3	0,0051	567	0,9676
11	10	19	0,0324	586	1,0000

Figure 16

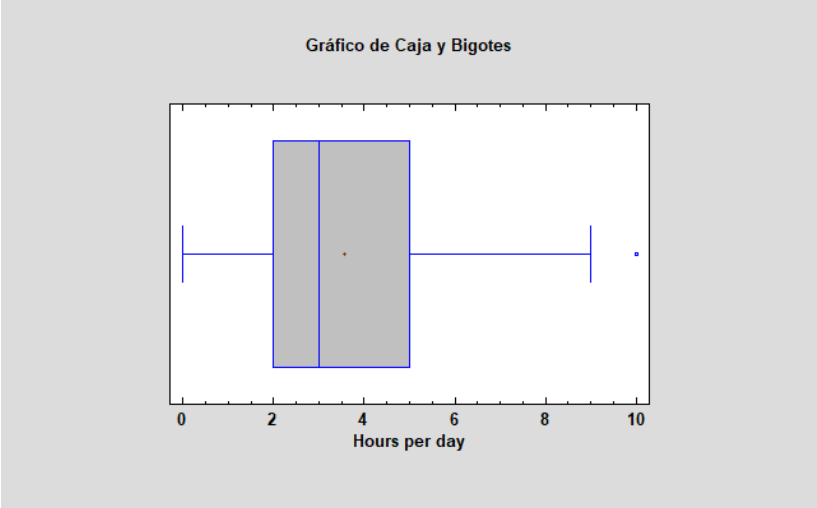

Figure 17

As it can be seen in the Box and Whisker chart there are 19 atypical data in the sample, which value is 10 hours. Furthermore, to describe its form, the standardized asymmetry coefficient of Fisher of 10,14 indicates a great positive asymmetry in the variable. It can be seen better in the bar chart and the scatter plot.

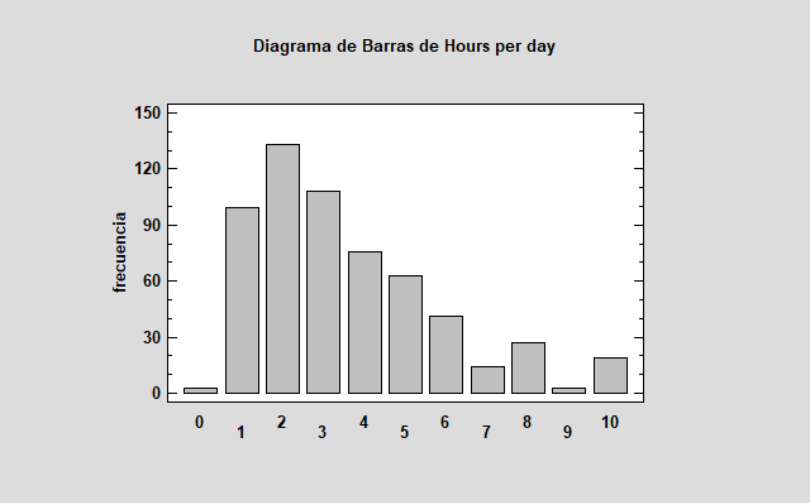

Figure 18

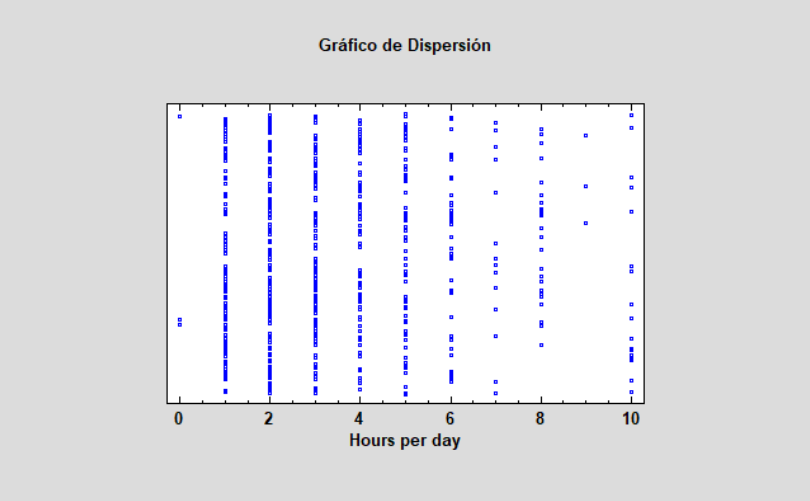

Figure 19

Regarding the position of the values, the median is 3, the lower quartile is 2 and the upper quartile is 5. Besides, about the dispersion of the sample, the interquartile range is 3.

While working

Indicates if each individual listens to music while he or she is working. It is a binary or dichotomous qualitative variable, since it only has two values: “Yes” or “No”.

Tabla de Frecuencia para While working

Clase	Valor	Frecuencia	Frecuencia Relativa	Frecuencia Acumulada	Frecuencia Rel. acum.
1	No	118	0,2014	118	0,2014
2	Yes	468	0,7986	586	1,0000

Figure 20

In the frequency table it can be observed that 468 individuals of 586 listen to music while they are working, a much higher value than the 118 that do not. The individuals that listen to music working are a 79,86% of the sample, while the ones who do not do it are a 20,14%. It is a thing easily appreciable in the bar chart and the pie chart.

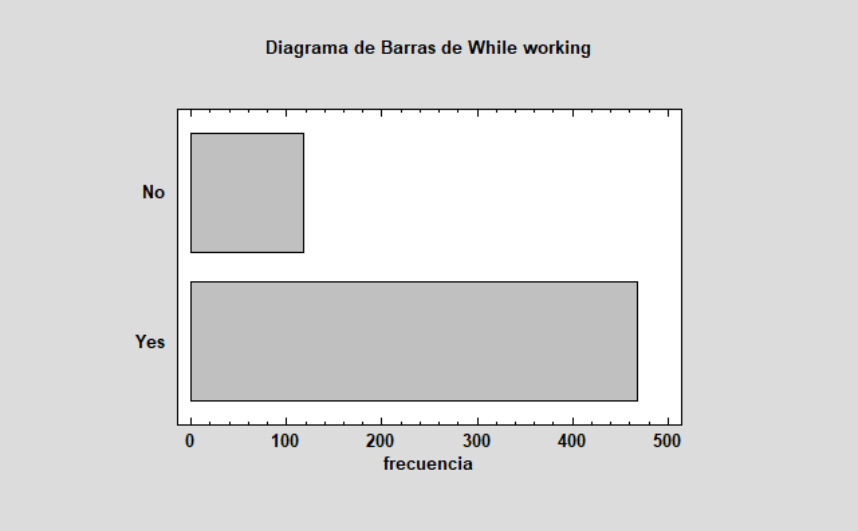

Figure 21

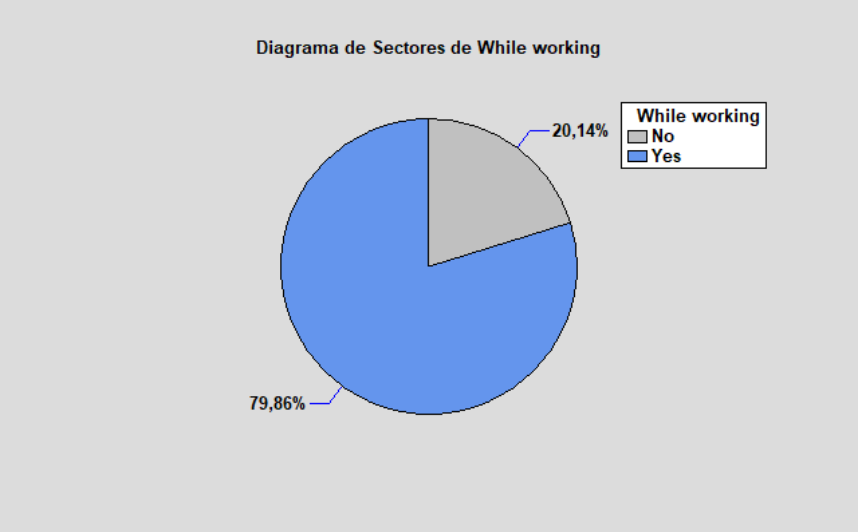

Figure 22

Exploratory

In this variable is indicated if each individual is willing to search and explore new types and genres of music. It is a binary or dichotomous qualitative variable, since it only has two values: "Yes" or "No".

Tabla de Frecuencia para Exploratory

Clase	Valor	Frecuencia	Frecuencia Relativa	Frecuencia Acumulada	Frecuencia Rel. acum.
1	No	157	0,2679	157	0,2679
2	Yes	429	0,7321	586	1,0000

Figure 23

In the frequency table it can be seen that the 73,21% of the individuals of the sample are willing to explore new music, instead, 26,79% are not willing to do it. The difference between both is observed in a better way in the bar chart and the pie chart.

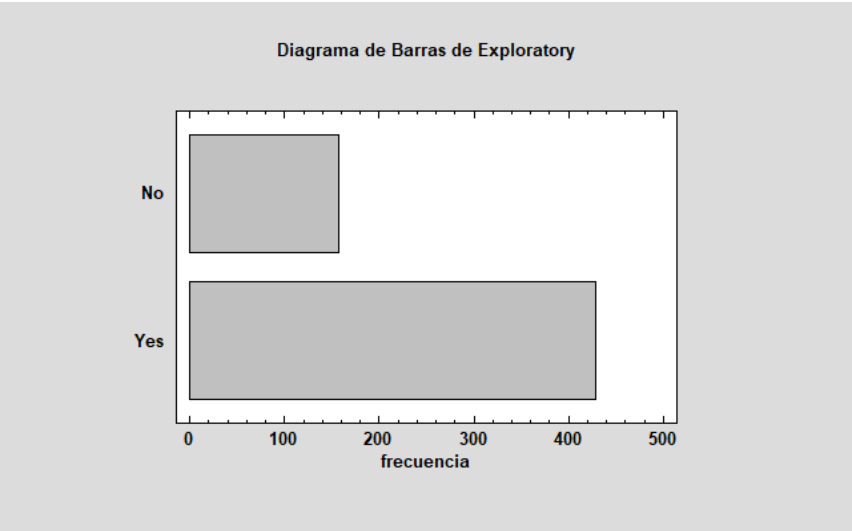

Figure 24

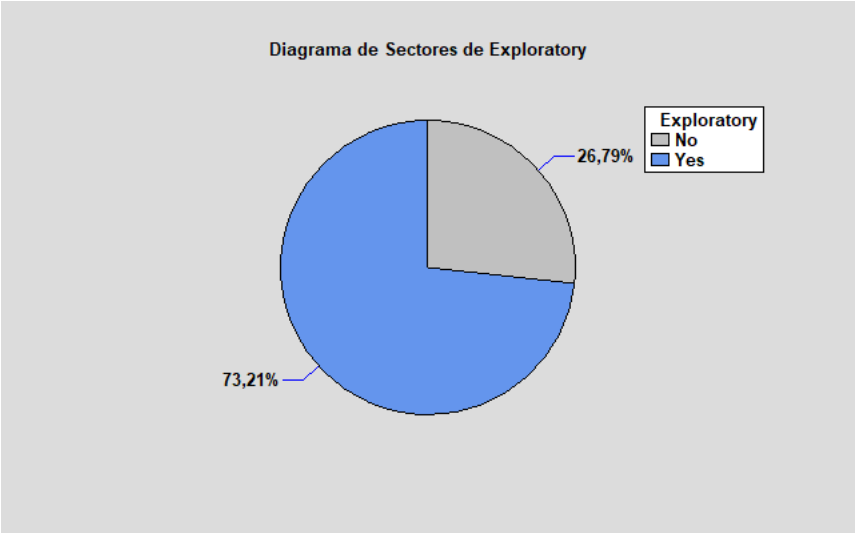

Figure 25

Foreign languages

Here it is shown if the individuals listen to music in a different language of their mother tongue. It is a binary or dichotomous qualitative variable, since it only has two values: “Yes” or “No”.

Tabla de Frecuencia para Foreign languages

Clase	Valor	Frecuencia	Frecuencia Relativa	Frecuencia Acumulada	Frecuencia Rel. acum.
1	No	255	0,4352	255	0,4352
2	Yes	331	0,5648	586	1,0000

Figure 26

In the frequency table as in the bar chart and the pie chart, it can be appreciated that both values are quite even. A 56,48% of the individuals listen to music in a foreign language, while a 43,52% do not.

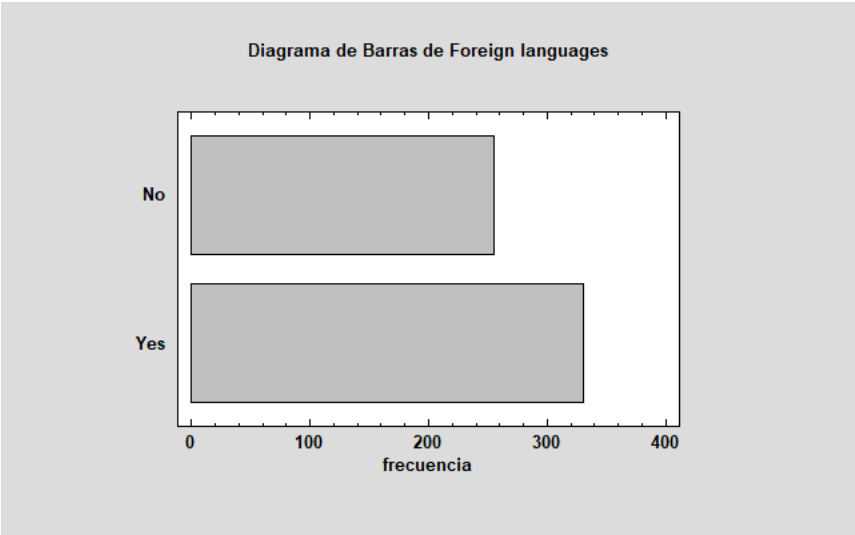

Figure 27

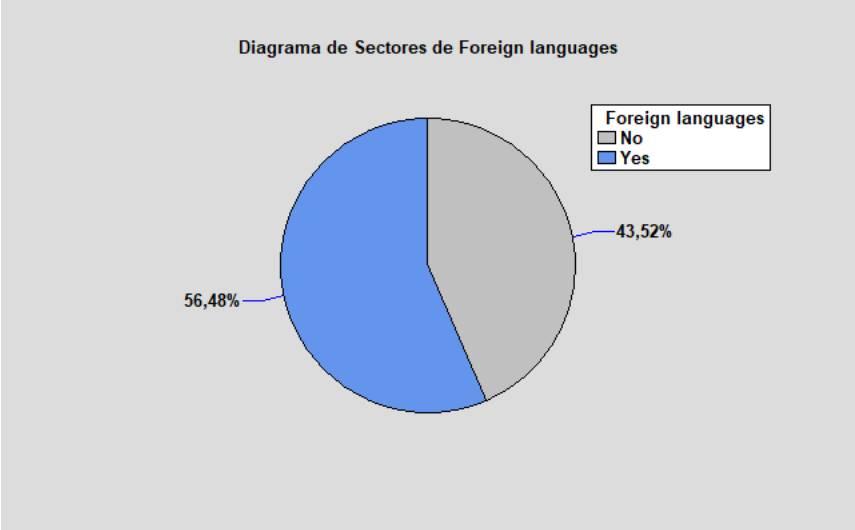

Figure 28

Composer

In this variable it can be seen if each individual of the sample composes or has composed any type of music. It is a binary or dichotomous qualitative variable, because it only has two values: "Yes" or "No".

Tabla de Frecuencia para Composer

Clase	Valor	Frecuencia	Frecuencia Relativa	Frecuencia Acumulada	Frecuencia Rel. acum.
1	No	489	0,8345	489	0,8345
2	Yes	97	0,1655	586	1,0000

Figure 29

In the frequency table the existence of a great difference between both values is appreciated. Between all the individuals, 489, a 83,45%, have never composed anything; while 97, a 16,55%, have composed something. This can be observed better in the bar chart and pie chart.

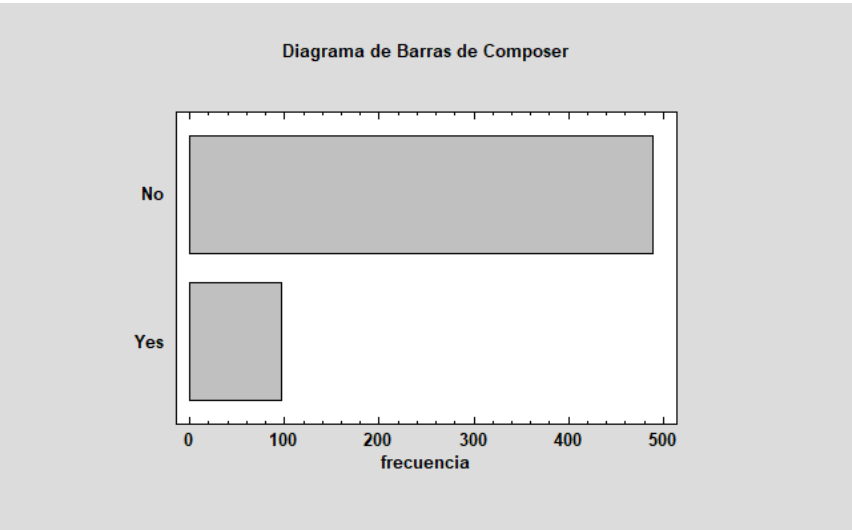

Figure 30

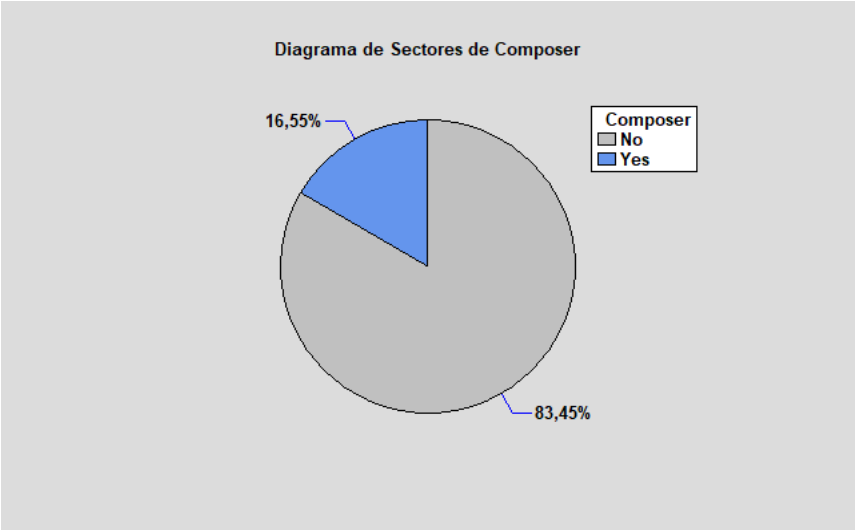

Figure 31

Music effects

This variable shows if the individuals feel that music improves or not their mood. It is an ordinal qualitative variable, due to the fact that the values follow an order. The values of the variable, ordered from best to worst, are: “Improve”, “No effect” and “Worsen”.

Tabla de Frecuencia para Music effects

Clase	Valor	Frecuencia	Frecuencia Relativa	Frecuencia Acumulada	Frecuencia Rel. acum.
1	Improve	446	0,7611	446	0,7611
2	No effect	127	0,2167	573	0,9778
3	Worsen	13	0,0222	586	1,0000

Figure 32

In the frequency table it is seen that a 76,11% of the individuals of the sample think that music improves their mood, while a 2,22% think that it worsens their mood. Moreover, music does not affect to a 21,67% of the individuals in their mood. Between them, highlights that music affect to the mood with 446 of the people; this is better appreciable in the bar chart and pie chart.

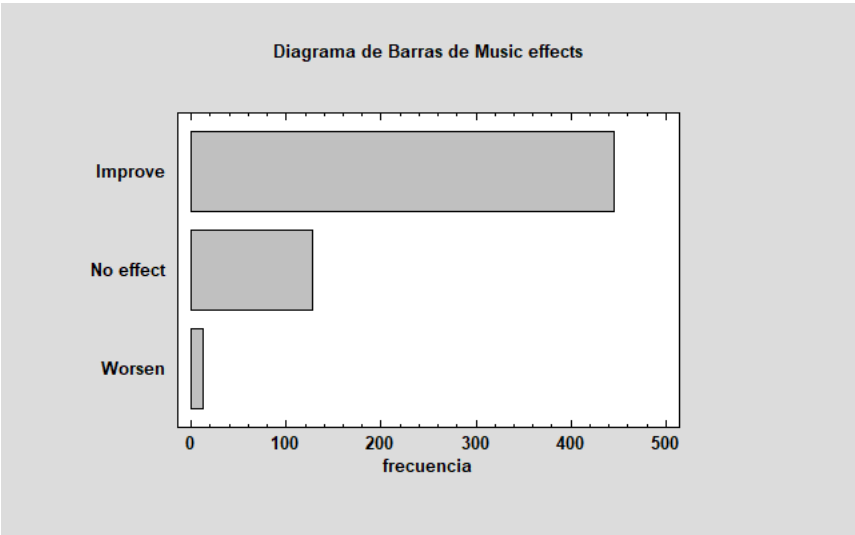

Figure 33

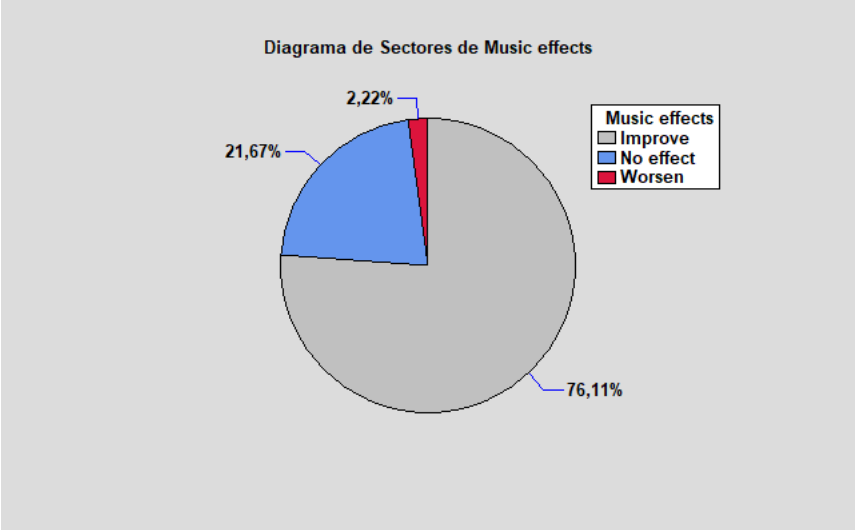

Figure 34

Fav genre

This variable shows the favourite genre of music of each individual of the sample. It is a nominal qualitative variable, and its values are: “Classical”, “Country”, “EDM”, “Folk”, “Gospel”, “Hip hop”, “Jazz”, “K pop”, “Latin”, “Lofi”, “Metal”, “Pop”, “R&B”, “Rap”, “Rock” and “Video game music”.

Tabla de Frecuencia para Fav genre

Clase	Valor	Frecuencia	Frecuencia Relativa	Frecuencia Acumulada	Frecuencia Rel. acum.
1	Classical	36	0,0614	36	0,0614
2	Country	21	0,0358	57	0,0973
3	EDM	34	0,0580	91	0,1553
4	Folk	23	0,0392	114	0,1945
5	Gospel	4	0,0068	118	0,2014
6	Hip hop	29	0,0495	147	0,2509
7	Jazz	16	0,0273	163	0,2782
8	K pop	23	0,0392	186	0,3174
9	Latin	2	0,0034	188	0,3208
10	Lofi	10	0,0171	198	0,3379
11	Metal	75	0,1280	273	0,4659
12	Pop	94	0,1604	367	0,6263
13	R&B	30	0,0512	397	0,6775
14	Rap	18	0,0307	415	0,7082
15	Rock	138	0,2355	553	0,9437
16	Video game music	33	0,0563	586	1,0000

Figure 35

As it is seen in the frequency table, the three more chosen genres of music as favourite for the individuals are: “Rock”, with a 23,55% of people of the sample; “Pop”, with a 16,04%; and “Metal”, with a 12,8%. On the other hand, the three ones less chosen are: “Lofi”, with a 1,71% of the individuals; “Gospel”, with a 0,68%; and “Latin”, with a 0,34%. That can be seen more visually in the bar chart and the pie chart.

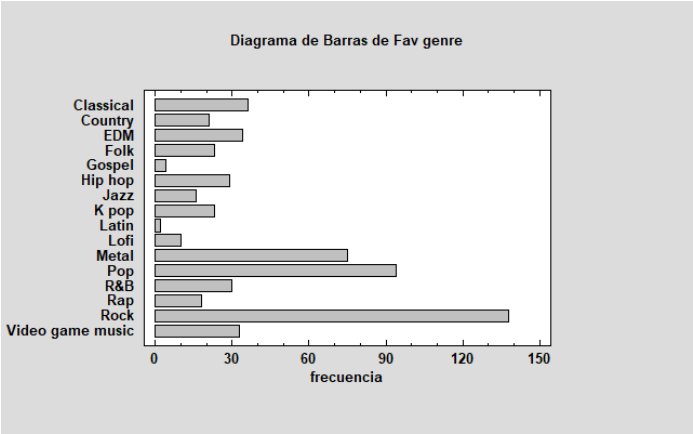

Figure 36

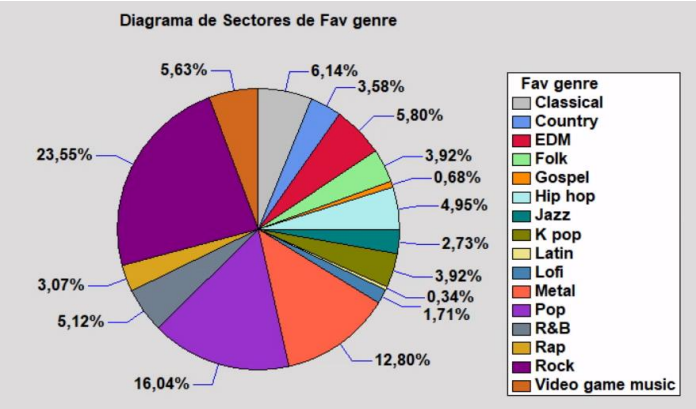

Figure 37

BPM

This variable shows the beats per minute in the favourite genre of every individual in the sample. It is a continuous quantitative variable, and its values range from 40 to 220 bpm.

This variable has been discretized in intervals of 40 in order to do the frequency table and the bar chart. Here are both:

Tabla de Frecuencia para BPM_Recod

Clase	*Valor*	*Frecuencia*	*Frecuencia Relativa*	*Frecuencia Acumulada*	*Frecuencia Rel. acum.*
1	B [40-80)	39	0,0666	39	0,0666
2	C [80-120)	221	0,3771	260	0,4437
3	D [120-160)	240	0,4096	500	0,8532
4	E [160-200)	73	0,1246	573	0,9778
5	F [200-240)	13	0,0222	586	1,0000

Figure 38

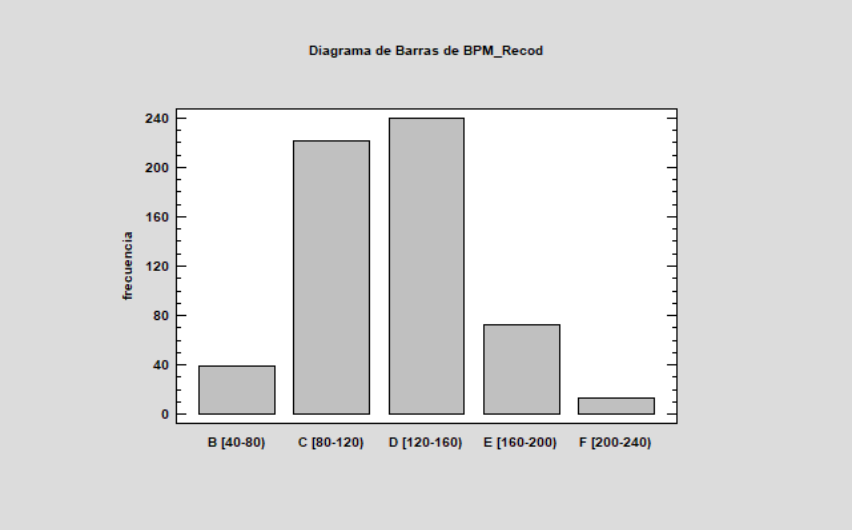

Figure 39

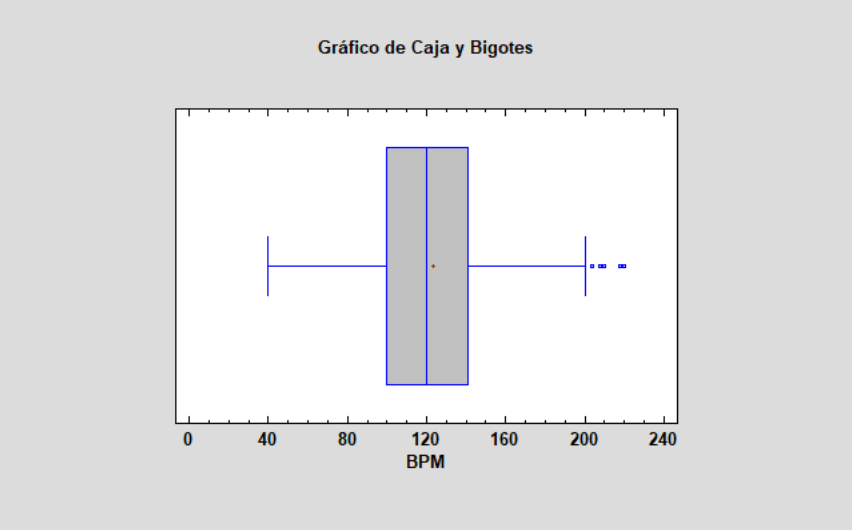

Figure 40

As it can be seen in the Box and Whisker chart there are 13 atypical data in the sample, which values go from 204 to 220 bpm. Furthermore, to describe its form, the standardized asymmetry coefficient of Fisher of 3,96 indicates a positive asymmetry in the variable. Moreover, it can be seen in a more visual way in the normal probability graph, the histogram and the scatter plot.

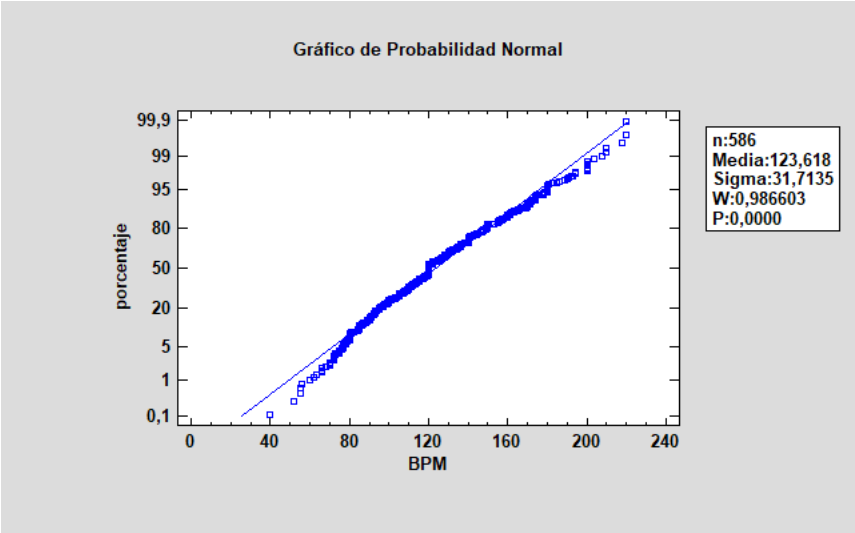

Figure 41

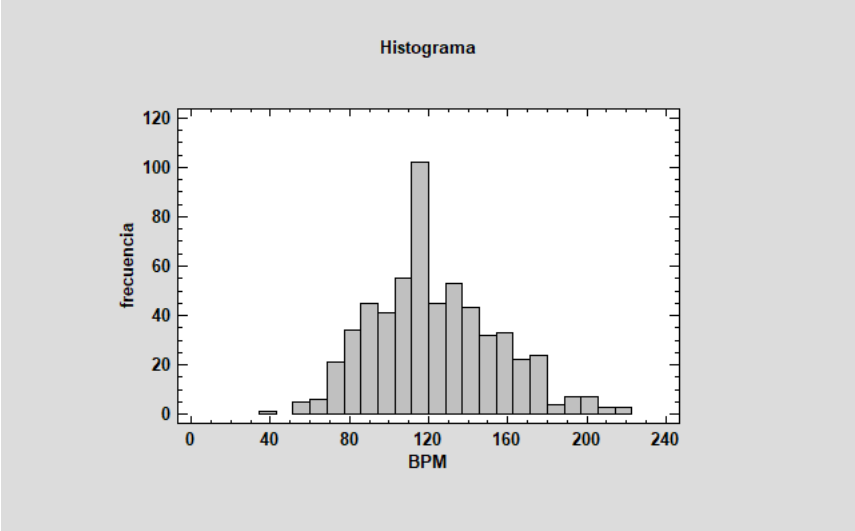

Figure 42

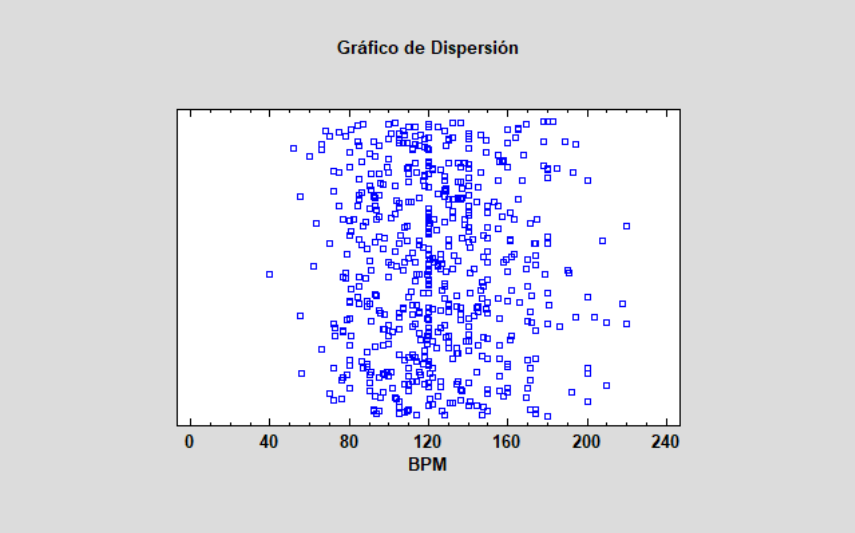

Figure 43

Regarding the position of the values, the median is 120, the lower quartile is 100 and the upper quartile is 141. Besides, about the dispersion of the sample, the interquartile range is 41.

Age

This variable shows the age of each individual in the sample. It is a continuous quantitative variable, and its values range from 10 to 80 years.

This variable has been discretized in intervals in order to do the frequency table and the bar chart. Here are both:

Tabla de Frecuencia para Age_Recod

Clase	Valor	Frecuencia	Frecuencia Relativa	Frecuencia Acumulada	Frecuencia Rel. acum.
1	(0-16]	75	0,1280	75	0,1280
2	(16-30]	402	0,6860	477	0,8140
3	(30-45]	72	0,1229	549	0,9369
4	(45-60]	24	0,0410	573	0,9778
5	(60-75]	12	0,0205	585	0,9983
6	(75-90]	1	0,0017	586	1,0000

Figure 44

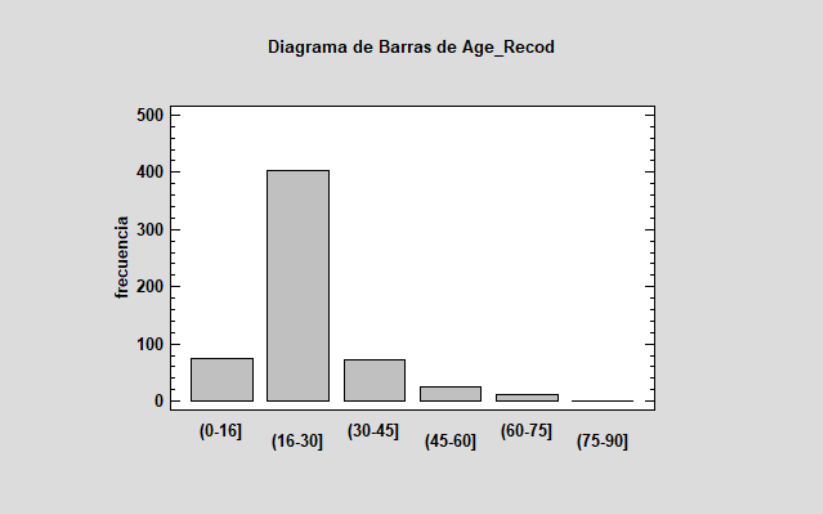

Figure 45

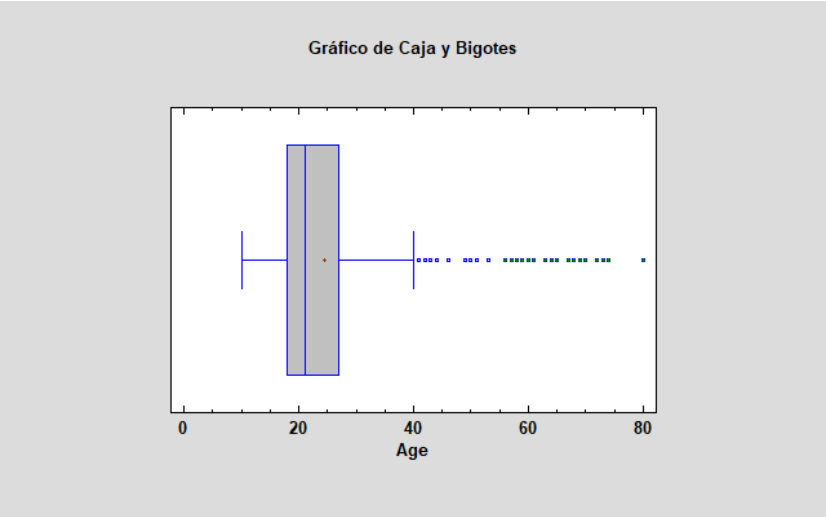

Figure 46

As it can be seen in the Box and Whisker chart there are 48 atypical data in the sample, which values go from 41 to 80 years. Furthermore, to describe its form, the standardized asymmetry coefficient of Fisher of 21,99 indicates a great positive asymmetry in the variable. Moreover, it can be seen in a more visual way in the normal probability graph, the histogram and the scatter plot.

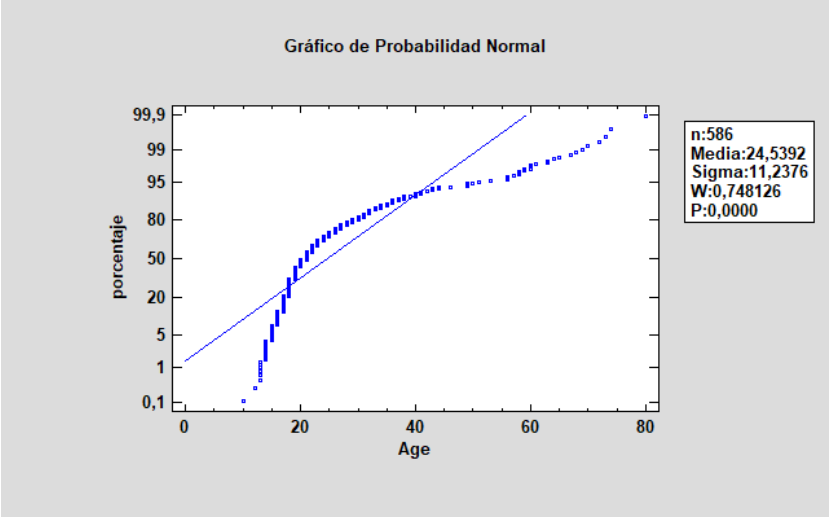

Figure 47

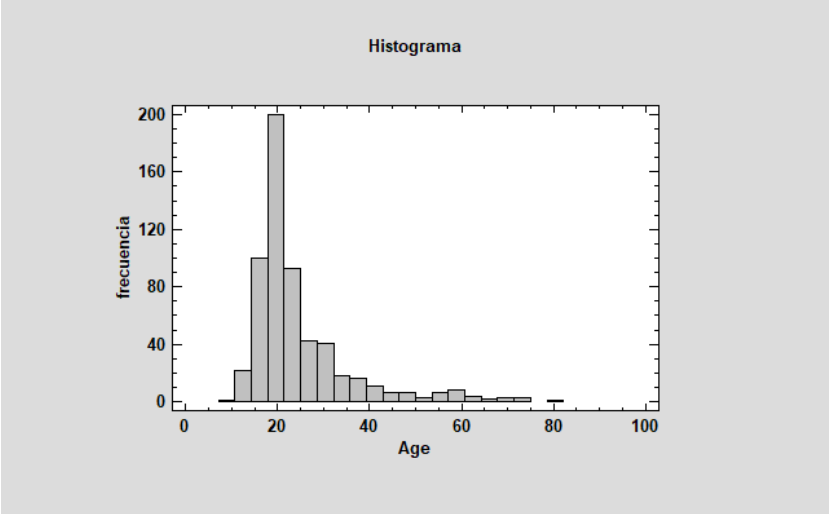

Figure 48

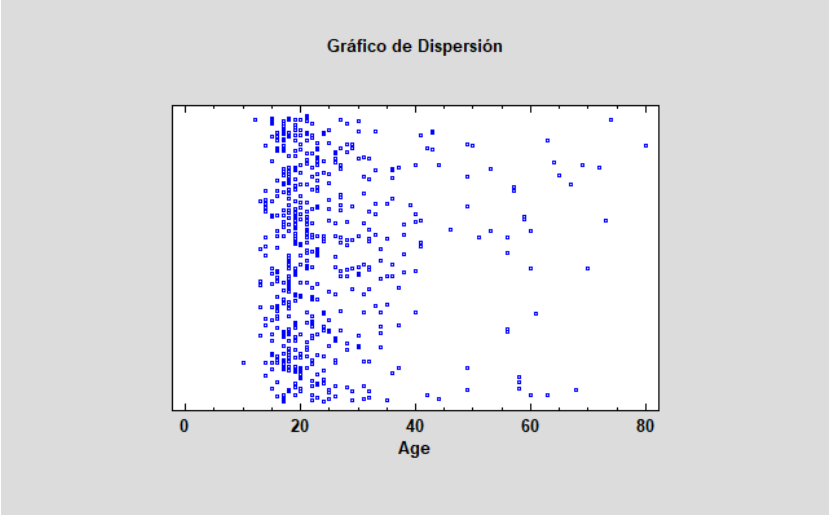

Figure 49

Regarding the position of the values, the median is 21, the lower quartile is 18 and the upper quartile is 27. Besides, about the dispersion of the sample, the interquartile range is 9.