

Equipo_9_DescriptiveStatistics

December 11, 2021

1 Descriptive Statistics: Measures of central tendency and Measures of dispersion

1.1 Team 9

Collaborators:

- Marcos Ambriz Flores
- Martha Patricia Chavez González
- Karina Becerril Rodríguez
- Armando Bringas Corpus

1.2 Problem Statement: El estrés en los estudiantes Psicología LX (EEP-10 application)

Instrument Application: <https://docs.google.com/forms/d/1sleQAccli8kbfaVLIZ2T2I5qlHg6j3FyZ8cXSa2XCaA/e>

Github Repository: <https://github.com/armandoBringas/JupyterNotebookStatistics>

WARNING: To render correctly Notebook please open it by executing *OpenJupyterNotebook.bat* file or open it in a Anaconda Prompt running the next line: *jupyter trust Equipo_9_DescriptiveStatistics.ipynb*

```
[1]: # See the "descriptiveStatistics.py" file to check the functions to get the ↵  
      ↳ descriptive statistics
```

```
import descriptiveStatistics as ds
```

```
[2]: # Read ".csv" to convert to a dataframe
```

```
df = ds.pd.read_csv('data.csv')
```

2 Data

```
[3]: display(df)
```

	Sexo	Edad	Nivel de estudios	Escolaridad	Cuatrimestre Cursando	(n) de materias cursando	R1
0	0	32	2	21	4	2	2
1	1	44	4	20	5	3	1
2	1	36	0	17	5	2	2
3	0	29	4	20	2	3	0
4	1	31	4	20	2	3	0
5	1	31	0	17	2	3	0
6	1	24	4	20	4	2	2
7	1	48	1	28	2	3	2
8	1	62	4	25	2	3	2
9	0	25	4	20	2	3	2
10	1	27	4	20	2	3	2
11	1	36	3	15	2	3	0
12	1	45	2	20	5	3	2
13	1	36	3	16	4	3	1
14	1	38	4	18	0	3	0
15	1	34	4	17	2	3	2
16	1	33	4	18	2	3	0
17	1	40	3	15	2	3	0
18	1	27	2	20	1	1	2
19	1	23	3	15	2	3	0
20	1	19	3	15	2	3	0
21	0	30	4	20	2	2	2
22	1	29	4	19	2	3	1
23	1	25	3	16	2	3	2
24	1	20	3	16	2	3	0
25	0	32	4	18	2	3	0
26	1	18	3	14	2	3	0
27	0	21	3	15	3	2	2
28	1	32	4	17	2	3	2
29	1	21	4	20	2	3	0
30	1	27	4	19	2	3	2
31	0	26	4	16	2	3	0
32	1	36	4	19	3	3	0
33	0	44	3	12	4	3	2
34	0	39	3	12	2	3	1
35	1	41	0	9	2	3	1
36	1	28	4	17	2	3	2
37	1	21	3	14	2	3	0
38	1	21	4	19	4	2	2
39	0	24	4	20	0	2	2
40	1	34	4	19	1	3	2
41	0	22	3	18	2	3	2
42	1	25	4	21	2	3	2
43	1	22	3	19	2	3	2

```
[4]: # Variables definition (VI, VD, VA), get dataframe header list and slice the
      ↪ list to get the lists
      # of the independent and dependent variables

      header_names = list(df)
      independent_variables = header_names[0:5] + header_names[6:len(header_names) -
      ↪ 1]
      intervening_variable = header_names[5]
      dependent_variable = header_names[len(header_names) - 1]
```

2.1 Variables Definition

2.1.1 Independent Variables

```
[5]: display(ds.Markdown(", ".join(independent_variables)))
```

Sexo, Edad, Nivel de estudios, Escolaridad, Cuatrimestre Cursando, R1, R2, R3, R4, R5, R6, R7, R8, R9, R10, R11, R12

2.1.2 Dependent Variable

```
[6]: display(ds.Markdown(dependent_variable))
```

TOTAL

2.1.3 Intervening Variable

```
[7]: display(ds.Markdown(intervening_variable))
```

(n) de materias cursando

3 Measures of central tendency

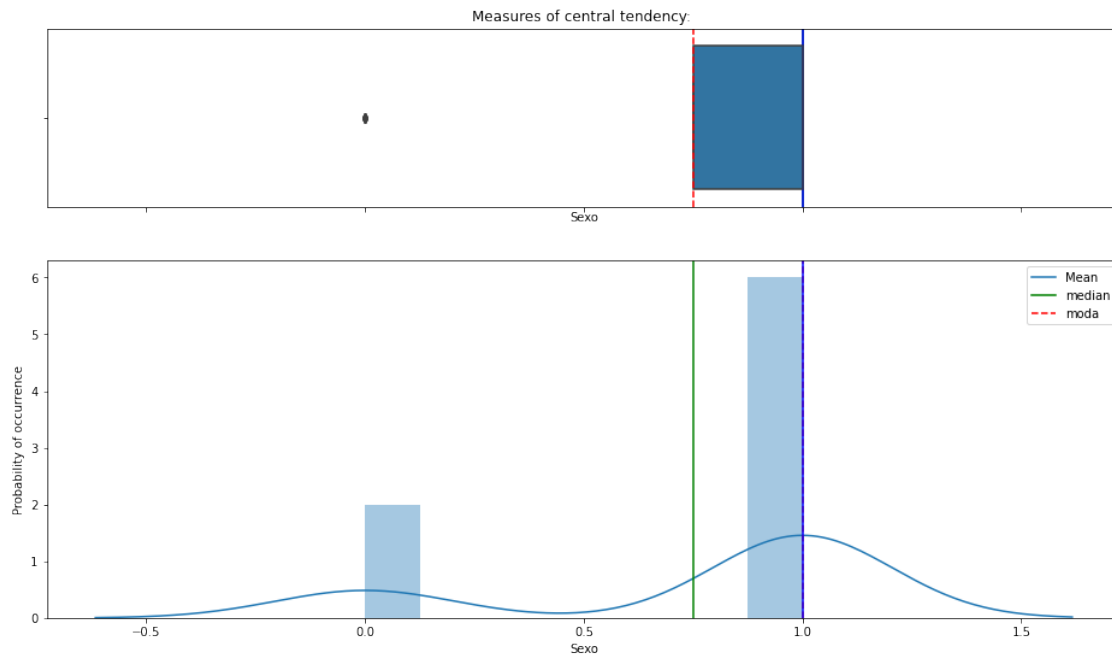
3.1 Independent Variables

```
[8]: # Calculate measures of central tendency for the Independent Variables

      for independent_variable in independent_variables:
          ds.measures_of_center(df, independent_variable)
```

3.1.1 Sexo

	Nominal Value	Categorical Value
0	Hombre	0
1	Mujer	1



Mean: 0.75

Median: 1.0

Mode: 1

Analysis:

Kurtosis: -0.6667, Platykurtic

Skewness: -1.1547, Left Asymmetry (+)

Normality Test:

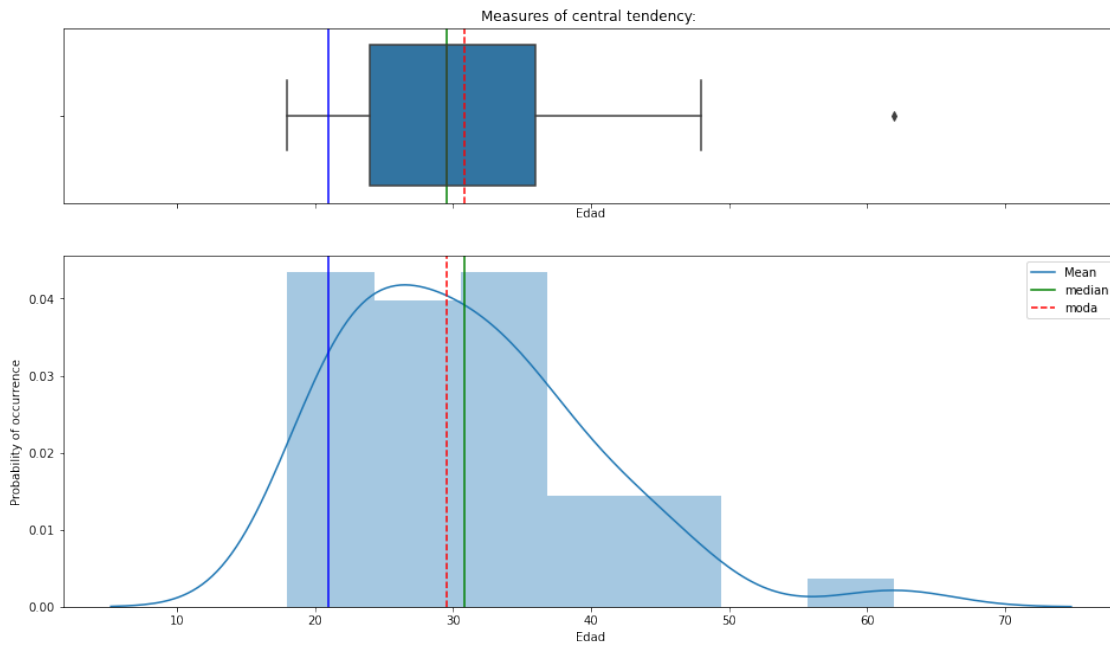
Saphiro-Wilk Test: $W = 0.5392$ $p\text{-value} = 1.499738389032501e-10$

H1: Variable has a non-normal distribution

Interpretation of Data:

From genre data, there is more women (75%) than men (25%).

3.1.2 Edad



Mean: 30.8636

Median: 29.5

Mode: 21

Analysis:

Kurtosis: 1.4375, Leptokurtic

Skewness: 1.0473, Right Assymetry (+)

Normality Test:

Saphiro-Wilk Test: $W = 0.9294$ $p\text{-value} = 0.009914972819387913$

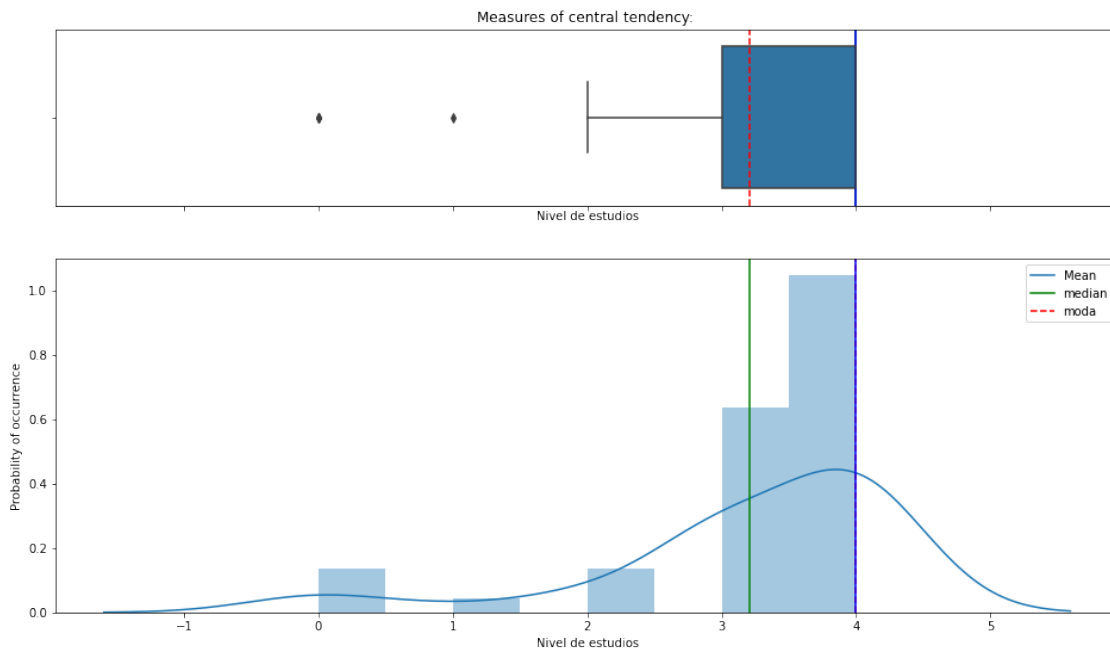
H1: Variable has a non-normal distribution

Interpretation of Data:

From age data, mean and median have approximately the same value however the mode is skewed. We can observe from histogram that ages were concentrated in 20-35 years range and we have an outlier that correspond from someone of 62 years, that causes to have the skewness and the mean and median distanced from mode.

3.1.3 Nivel de estudios

Nominal Value	Categorical Value
0 Carrera técnica con preparatoria terminada	0
1 Doctorado	1
2 Maestría	2
3 Preparatoria / Bachillerato	3
4 Profesional	4



Mean: 3.2045

Median: 4.0

Mode: 4

Analysis:

Kurtosis: 2.1416, Leptokurtic

Skewness: -1.6714, Left Assymetry (+)

Normality Test:

Saphiro-Wilk Test: $W = 0.7015$ $p\text{-value} = 3.728138864289576e-08$

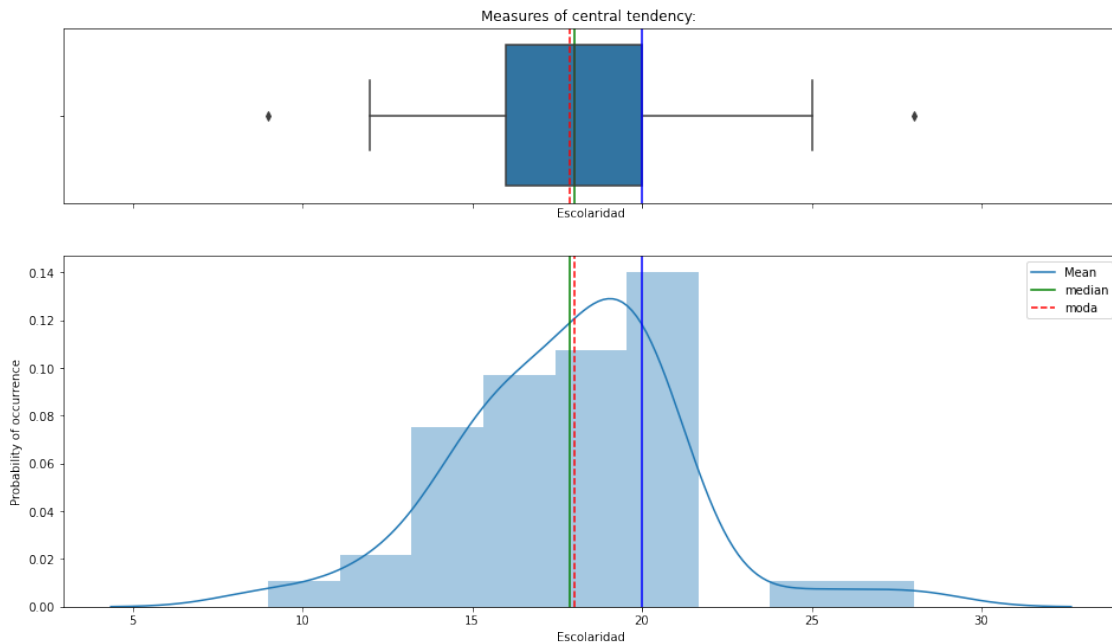
H1: Variable has a non-normal distribution

Interpretation of Data:

From studies level data, mean and mode have the same value however the mean is slightly skewed. We can observe from histogram that the studies level is

concentrated in 'Preparatoria' and 'Profesional' level. However, the reason why we don't have a normal is distribution is due that we have two outliers that corresponds to 'Carrera técnica con preparatoria terminada' and 'Doctorado'.

3.1.4 Escolaridad



Mean: 17.8636

Median: 18.0

Mode: 20

Analysis:

Kurtosis: 1.6271, Leptokurtic

Skewness: 0.1453, Right Assymetry (+)

Normality Test:

Saphiro-Wilk Test: $W = 0.9391$ $p\text{-value} = 0.021855933591723442$

H1: Variable has a non-normal distribution

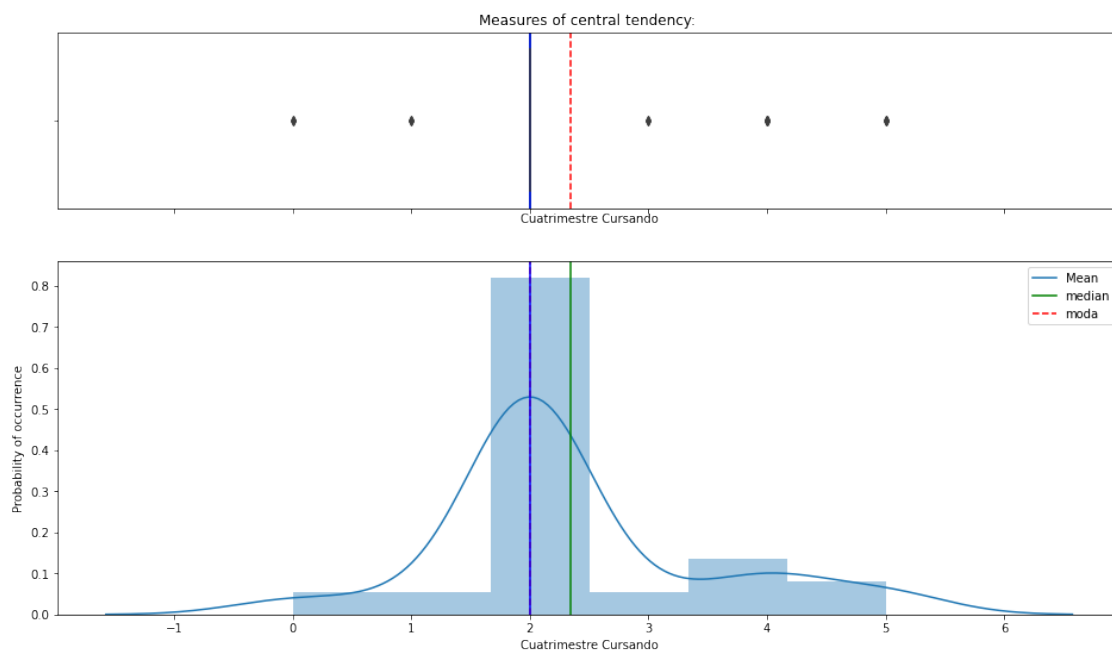
Interpretation of Data:

From years of accumulated study data, mean and median have the same value approximately the same value however the mode is skewed, data is concentrated around the 17-10 years of study. However, the reason why we don't have a normal

is distribution is due that we have two outliers that corresponds to lowest value 9 years and the highest value 28.

3.1.5 Cuatrimestre Cursando

	Nominal Value	Categorical Value
0	Cuarto	0
1	Noveno	1
2	Primero	2
3	Quinto	3
4	Segundo	4
5	Tercero	5



Mean: 2.3409

Median: 2.0

Mode: 2

Analysis:

Kurtosis: 0.8667, Leptokurtic

Skewness: 0.8074, Right Assymetry (+)

Normality Test:

Saphiro-Wilk Test: $W = 0.7517$ $p\text{-value} = 3.0183051080712175e-07$

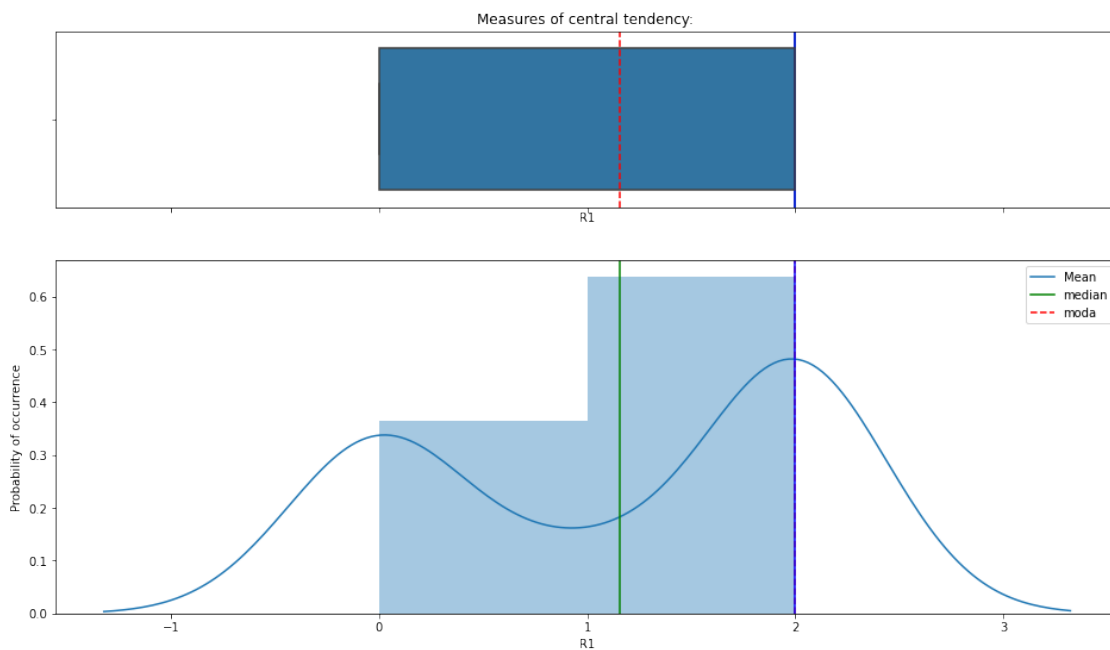
H1: Variable has a non-normal distribution

Interpretation of Data:

From quarter studying data, mean, median, mode have the same value approximately the same value, that corresponds from people that are in first quarter. However, the reason why we don't have a normal is distribution is due that we have five outliers that corresponds to people that are in other quarters, the dispersion of the data is high, just a slight concentration of people that is on the first quarter.

3.1.6 R1 : ¿Dónde percibes con mayor frecuencia estrés?

	Nominal Value	Categorical Value
0	Escuela	0
1	Hogar	1
2	Trabajo	2



Mean: 1.1591

Median: 2.0

Mode: 2

Analysis:

Kurtosis: -1.7621, Platykurtic

Skewness: -0.3203, Left Assymetry (+)

Normality Test:

Saphiro-Wilk Test: $W = 0.6957$ $p\text{-value} = 2.9717904581616494e-08$

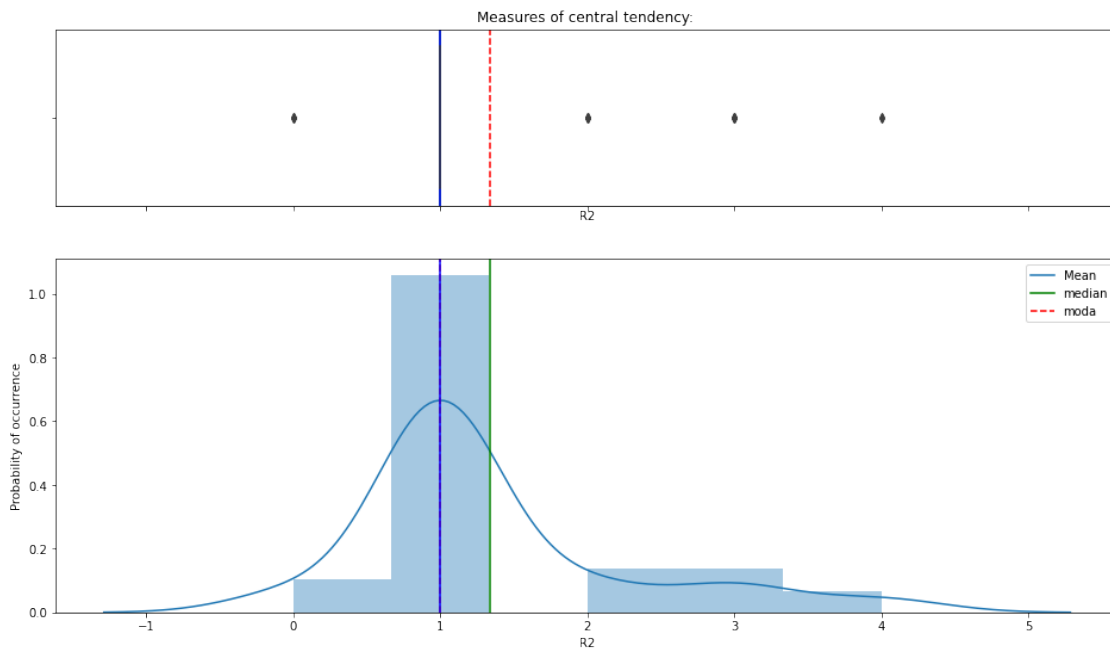
H1: Variable has a non-normal distribution

Interpretation of Data:

From R1, mean and mode have the same value however the mean is skewed. Mean and mode corresponds to 'Trabajo'. From data, percentages correspond to: 11.4% - 'Hogar', 36%.4 - 'Escuela' and 52.3% - 'Trabajo', half of the sample perceived in the work the stress.

3.1.7 R2 : En el ambiente universitario ¿Qué aspecto te provoca más estrés?

	Nominal Value	Categorical Value
0	Escuela	0
1	Hogar	1
2	Trabajo	2



Mean: 1.3409

Median: 1.0

Mode: 1

Analysis:

Kurtosis: 1.8232, Leptokurtic

Skewness: 1.4982, Right Assymetry (+)

Normality Test:

Saphiro-Wilk Test: $W = 0.6896$ $p\text{-value} = 2.3442499497150493e-08$

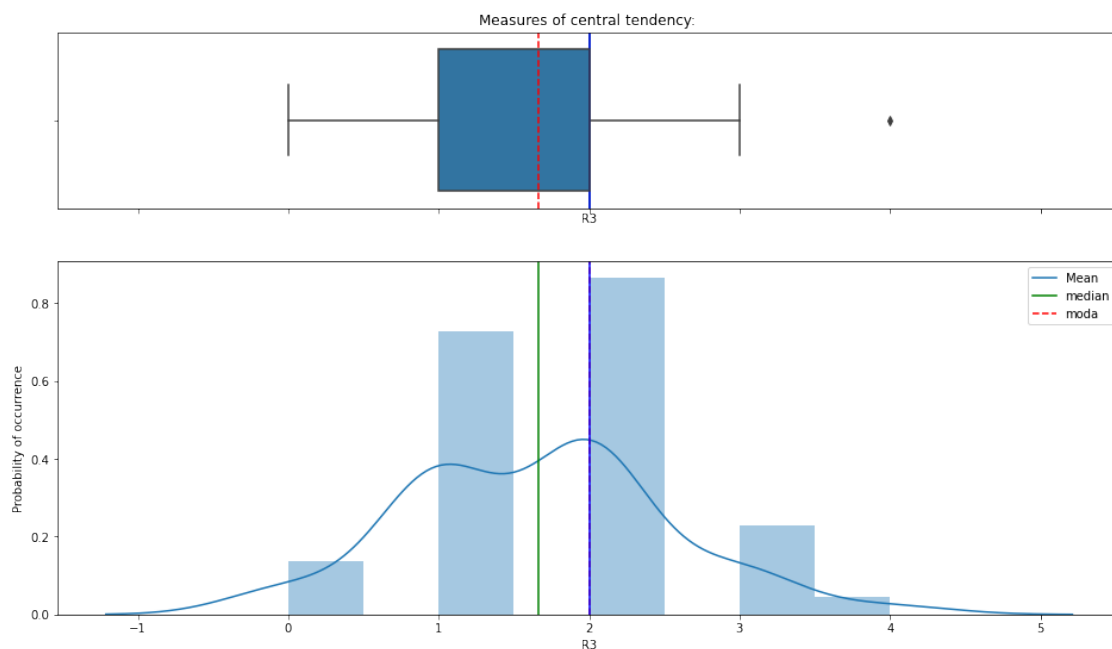
H1: Variable has a non-normal distribution

Interpretation of Data:

From R2, median and mode have the same value however the mean is skewed. Mean and mode corresponds to 'Profesores'. From the data 'Trabajo en Equipo' and 'Profesores' have the same percentage 9.1%. From the shape of the boxplot and distribution curve, we can observe that the data has a high dispersion among the responses, and we don't have a normal distribution due to the four outliers that appear in the data.

3.1.8 R3 : Respecto a las clases de la Licenciatura en Psicología ¿Con qué frecuencia has estado afectado por algo que ha ocurrido inesperadamente?

	Nominal Value	Categorical Value
0	Nunca	0
1	Casi nunca	1
2	De vez en cuando	2
3	A menudo	3
4	Muy a menudo	4



Mean: 1.6591

Median: 2.0

Mode: 2

Analysis:

Kurtosis: 0.1245, Leptokurtic

Skewness: 0.2709, Right Assymetry (+)

Normality Test:

Saphiro-Wilk Test: W = 0.8829 p-value = 0.00033702992368489504

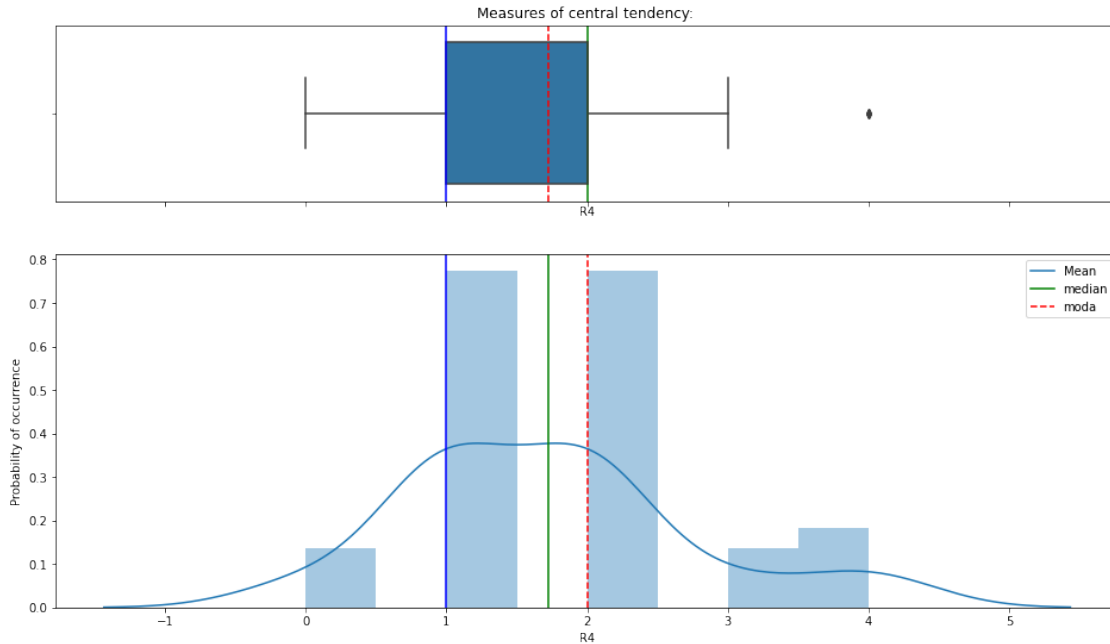
H1: Variable has a non-normal distribution

Interpretation of Data:

From R3, median and mode have the same value however the mean is skewed. Mean and mode corresponds to 'De vez en cuando'. Median corresponds to 75% percentile, 'De vez en cuando' percentage corresponds to 43.2%, there is an outlier that corresponds to 'Nunca'.

3.1.9 R4 : ¿Con qué frecuencia te has sentido incapaz de controlar las cosas importantes en tu vida?

	Nominal Value	Categorical Value
0	Nunca	0
1	Casi nunca	1
2	De vez en cuando	2
3	A menudo	3
4	Muy a menudo	4



Mean: 1.7273

Median: 2.0

Mode: 1

Analysis:

Kurtosis: 0.2144, Leptokurtic

Skewness: 0.6982, Right Assymetry (+)

Normality Test:

Saphiro-Wilk Test: $W = 0.8615$ $p\text{-value} = 8.688793604960665e-05$

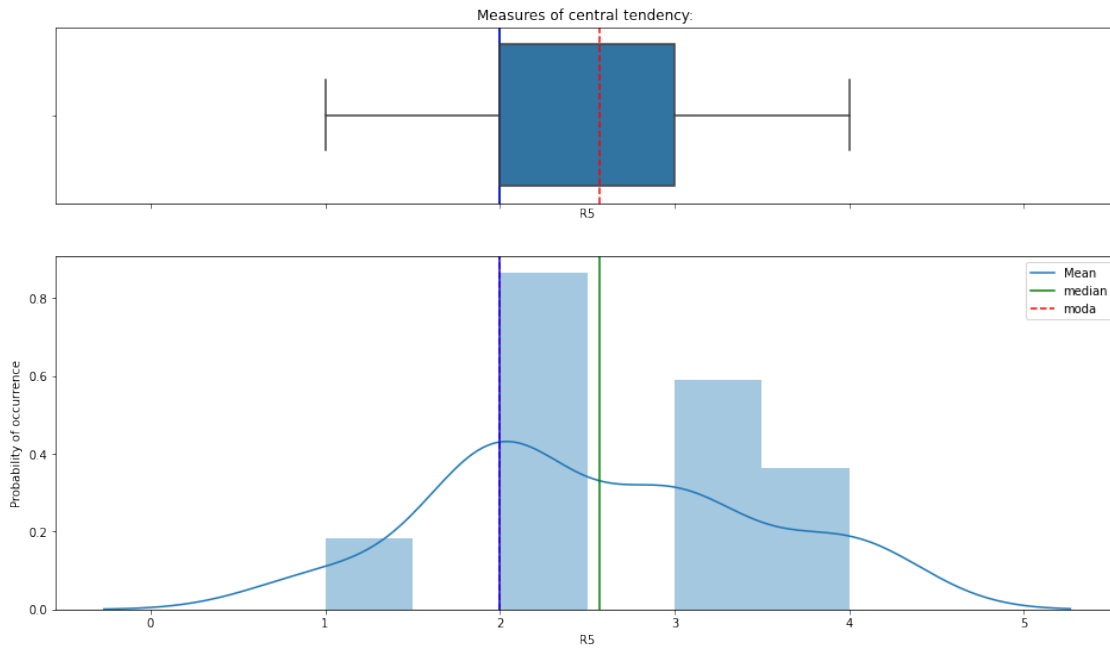
H1: Variable has a non-normal distribution

Interpretation of Data:

From R4, we have two Modes that correspond to 'Casi nunca' and 'De vez en cuando', both percentage values corresponds to 38.6%, that is the reason why we have a Leptokurtic shape due to the data concentration in 'Casi nunca' and 'De vez en cuando response'. There is an outlier that corresponds to 'Nunca'.

3.1.10 R5 : ¿Con qué frecuencia te has sentido nervioso o estresado?

	Nominal Value	Categorical Value
0	Nunca	0
1	Casi nunca	1
2	De vez en cuando	2
3	A menudo	3
4	Muy a menudo	4



Mean: 2.5682

Median: 2.0

Mode: 2

Analysis:

Kurtosis: -0.8105, Platykurtic

Skewness: 0.1816, Right Assymetry (+)

Normality Test:

Saphiro-Wilk Test: $W = 0.8648$ $p\text{-value} = 0.00010618809756124392$

H1: Variable has a non-normal distribution

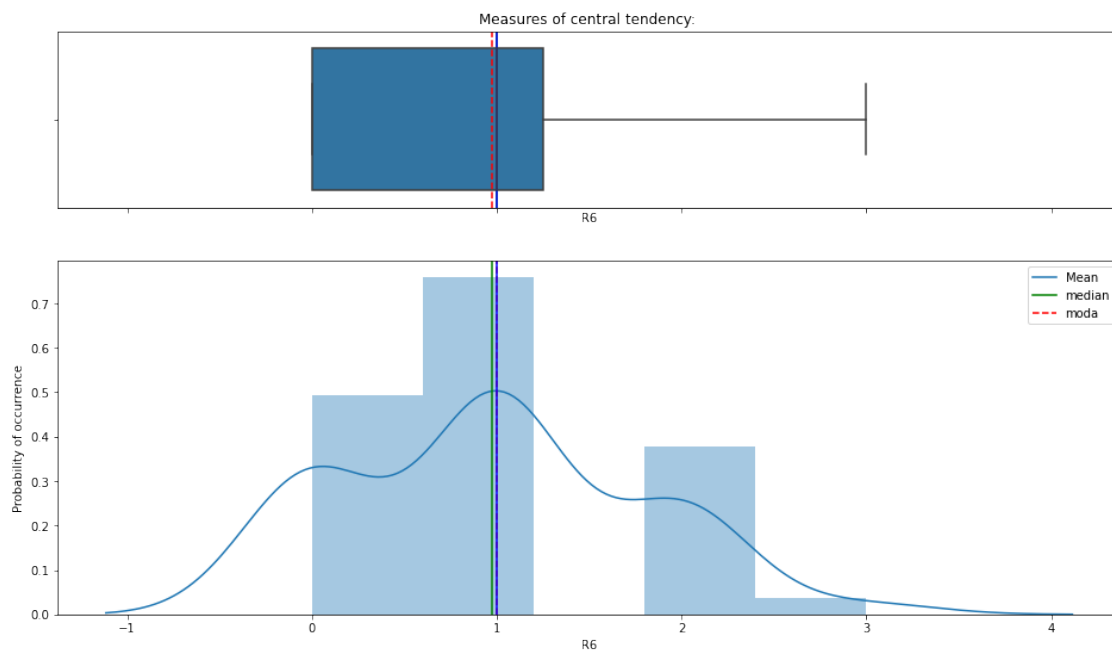
Interpretation of Data:

From R5, mean, median and mode have approximately the same value. The mean corresponds to 25% percentile that is the 'De vez en cuando' value and it has

the 43.2% percentage from the data. However, from the other responses there is a high dispersion. From this reactive, there wasn't responses from the sample for 'Nunca' option.

3.1.11 R6 : ¿Con qué frecuencia ha estado seguro sobre su capacidad para manejar sus problemas personales?

	Nominal Value	Categorical Value
0	Nunca	4
1	Casi nunca	3
2	De vez en cuando	2
3	A menudo	1
4	Muy a menudo	0



Mean: 0.9773

Median: 1.0

Mode: 1

Analysis:

Kurtosis: -0.6096, Platykurtic

Skewness: 0.3238, Right Assymetry (+)

Normality Test:

Saphiro-Wilk Test: $W = 0.8423$ $p\text{-value} = 2.798677633109037e-05$

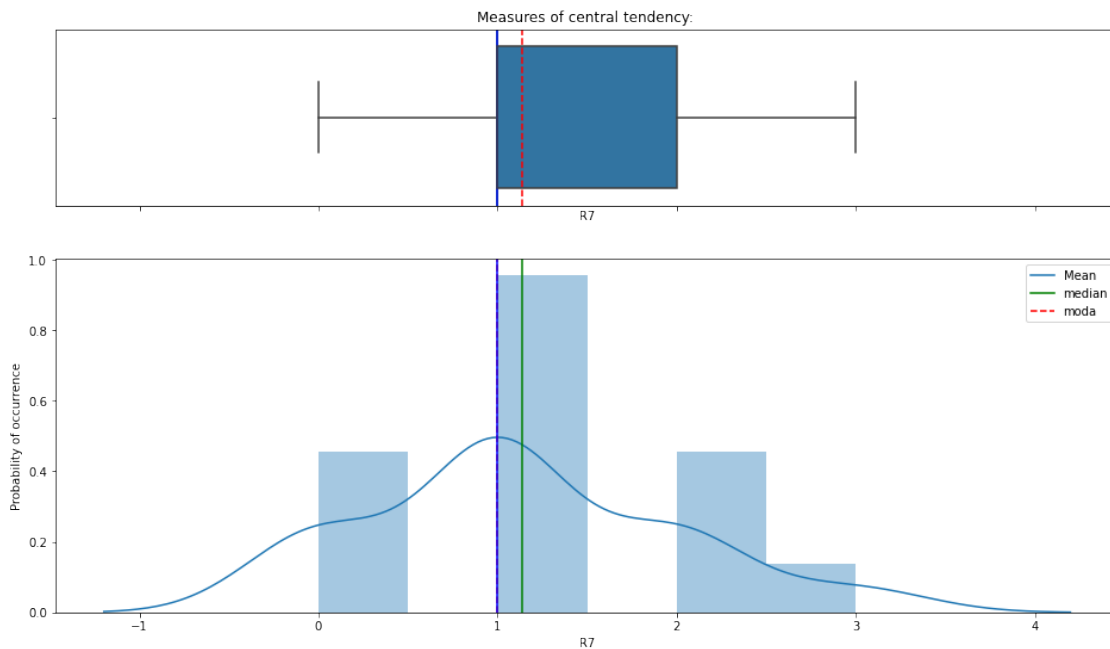
H1: Variable has a non-normal distribution

Interpretation of Data:

From R6, mean, median and mode have approximately the same value that corresponds to 'Casi nunca'. Observation from distribution shows that data are skewed and concentrated in the 'A menudo' - 45.5% percentage, half of the data, that is the reason why box blot is skewed to the first half, the boxplot starts in the first whisker.

3.1.12 R7 : ¿Con qué frecuencia has sentido que las cosas te van bien?

	Nominal Value	Categorical Value
0	Nunca	4
1	Casi nunca	3
2	De vez en cuando	2
3	A menudo	1
4	Muy a menudo	0



Mean: 1.1364

Median: 1.0

Mode: 1

Analysis:

Kurtosis: -0.3557, Platykurtic

Skewness: 0.4241, Right Assymetry (+)

Normality Test:

Saphiro-Wilk Test: $W = 0.8577$ $p\text{-value} = 6.904988549649715e-05$

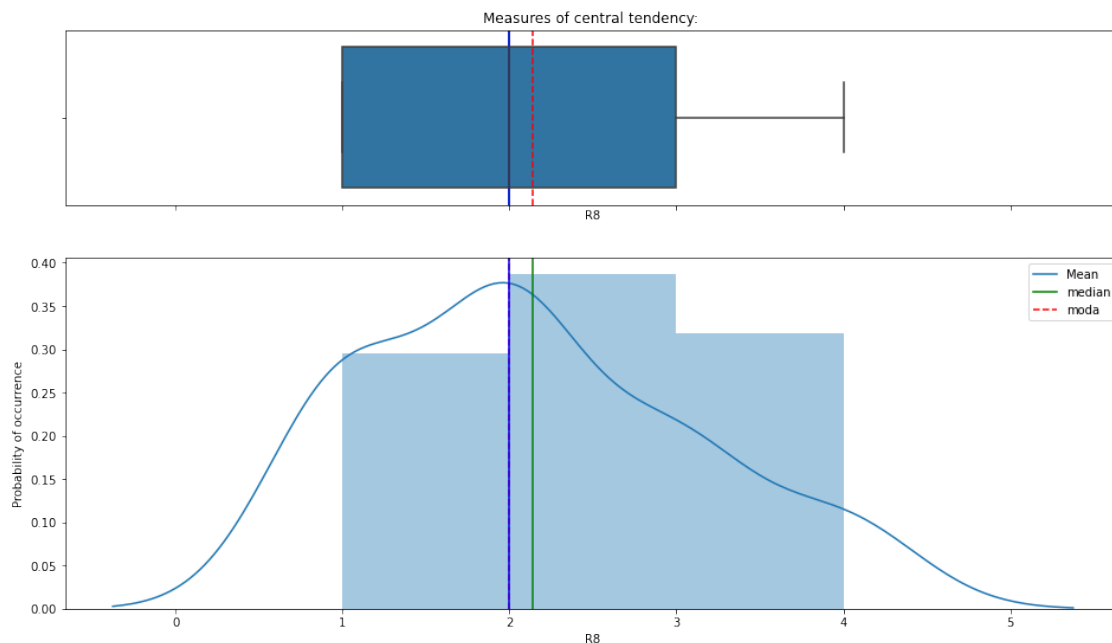
H1: Variable has a non-normal distribution

Interpretation of Data:

From R7, mean, median and mode have approximately the same value that corresponds to 'Casi nunca' that has 47.7%, mean and median corresponds to the 25% percentile that is the 'Casi nunca' value. From the other responses there is a high dispersion from the data.

3.1.13 R8 : ¿Con qué frecuencia has sentido que no podías afrontar todas las cosas que tenías que hacer?

	Nominal Value	Categorical Value
0	Nunca	0
1	Casi nunca	1
2	De vez en cuando	2
3	A menudo	3
4	Muy a menudo	4



Mean: 2.1364

Median: 2.0

Mode: 2

Analysis:

Kurtosis: -0.7427, Platykurtic

Skewness: 0.4779, Right Assymetry (+)

Normality Test:

Saphiro-Wilk Test: W = 0.8563 p-value = 6.362615386024117e-05

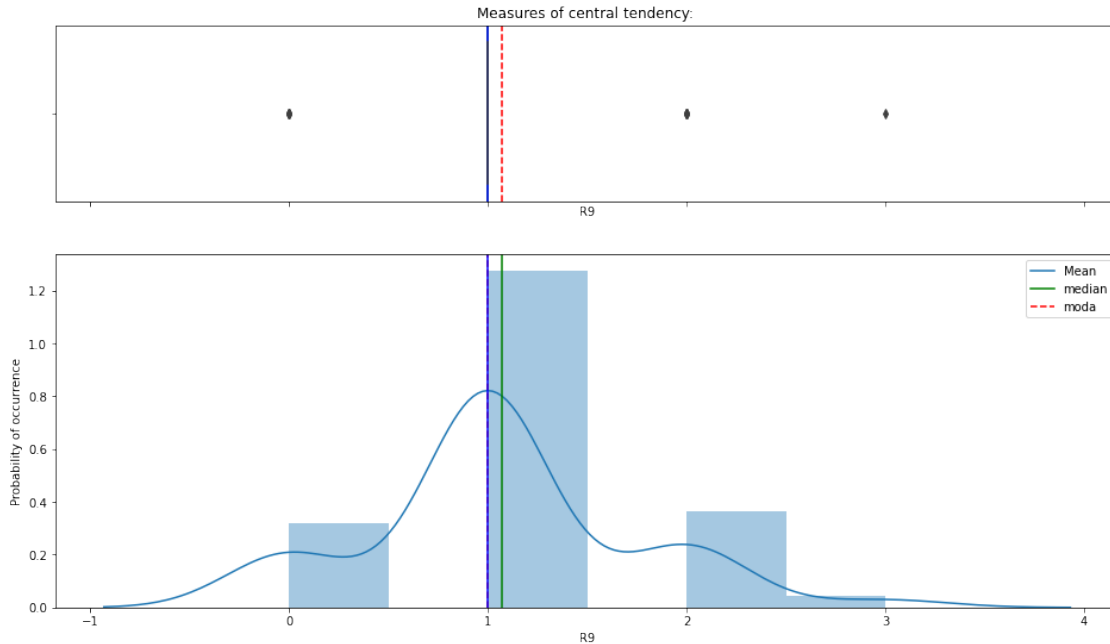
H1: Variable has a non-normal distribution

Interpretation of Data:

From R8, mean, median and mode have approximately the same value that corresponds to 'De vez en cuando'. Observation from distribution shows that data are skewed and concentrated in the 'De vez en cuando' - 38.6% percentage, approximately one third from the data, that is the reason why box blot is skewed to the first half, the boxplot starts in the first whisker.

3.1.14 R9 : ¿Con qué frecuencia has podido controlar las dificultades en tu vida?

	Nominal Value	Categorical Value
0	Nunca	4
1	Casi nunca	3
2	De vez en cuando	2
3	A menudo	1
4	Muy a menudo	0



Mean: 1.0682

Median: 1.0

Mode: 1

Analysis:

Kurtosis: 0.621, Leptokurtic

Skewness: 0.4185, Right Assymetry (+)

Normality Test:

Saphiro-Wilk Test: $W = 0.7928$ $p\text{-value} = 2.0403081180120353e-06$

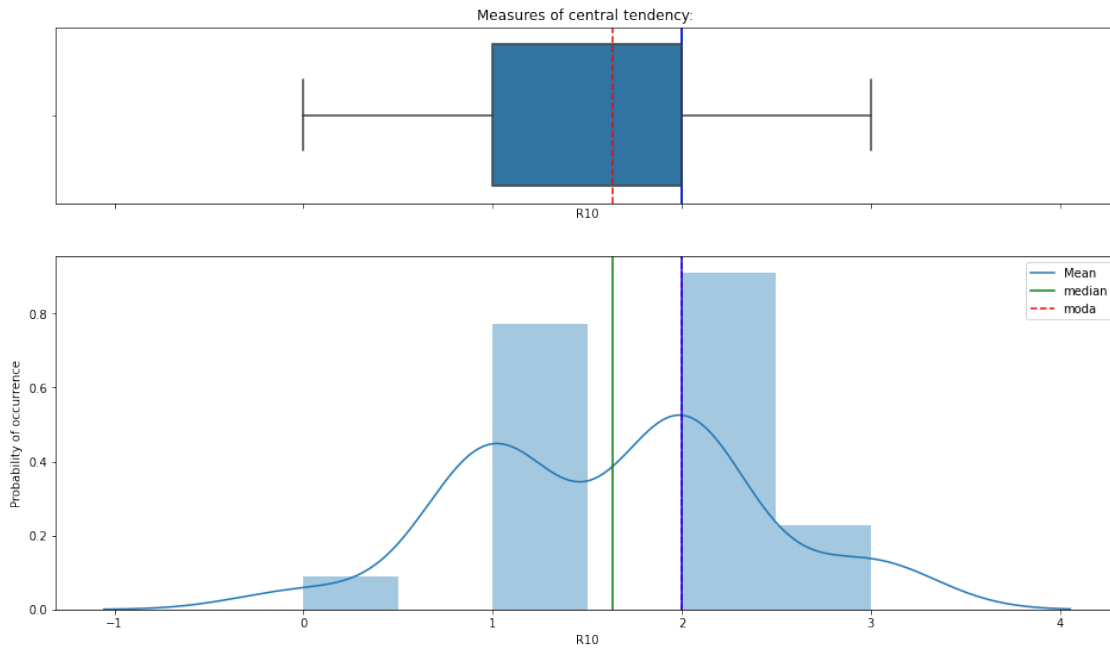
H1: Variable has a non-normal distribution

Interpretation of Data:

From R8, mean, median and mode have approximately the same value that corresponds to 'De vez en cuando' that corresponds to 63.6% percentage, that means two thirds from the data. The boxplot seems to be inexistent due to that the data is highly concentrated in 'De vez en cuando' value and the remaining data corresponds to the outliers.

3.1.15 R10 : ¿Con qué frecuencia te has sentido que tenía todo bajo control?

	Nominal Value	Categorical Value
0	Nunca	4
1	Casi nunca	3
2	De vez en cuando	2
3	A menudo	1
4	Muy a menudo	0



Mean: 1.6364

Median: 2.0

Mode: 2

Analysis:

Kurtosis: -0.384, Platykurtic

Skewness: 0.0277, Right Assymetry (+)

Normality Test:

Saphiro-Wilk Test: $W = 0.8485$ $p\text{-value} = 4.0101836930261925e-05$

H1: Variable has a non-normal distribution

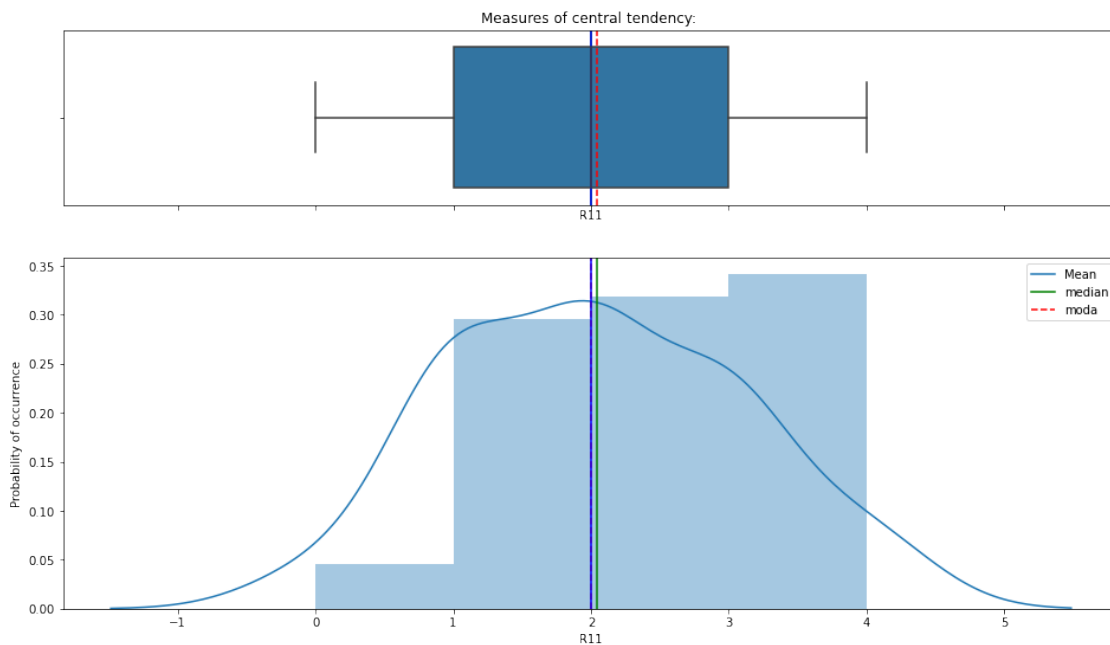
Interpretation of Data:

From R10, mean, median and mode have approximately the same value that corresponds to 'De vez en cuando'. Data is concentrated in 'De vez en cuando' -

45.5% and 'A menudo' - 38.6%. Median and mode corresponds to the 75% percentile and approximately match with the sum of the 45.5% and 38.6% values percentages.

3.1.16 R11 : ¿Con qué frecuencia has estado enfadado porque las cosas que te han ocurrido estaban fuera de tu control?

	Nominal Value	Categorical Value
0	Nunca	0
1	Casi nunca	1
2	De vez en cuando	2
3	A menudo	3
4	Muy a menudo	4



Mean: 2.0455

Median: 2.0

Mode: 2

Analysis:

Kurtosis: -0.737, Platykurtic

Skewness: 0.1493, Right Assymetry (+)

Normality Test:

Saphiro-Wilk Test: $W = 0.9089$ $p\text{-value} = 0.0020615924149751663$

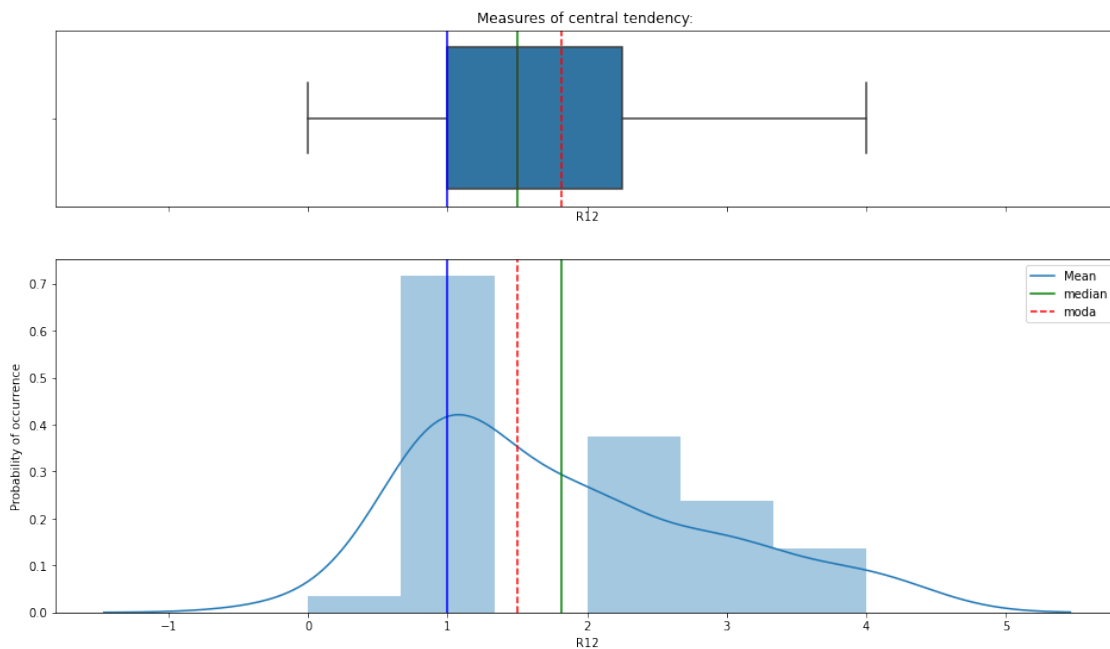
H1: Variable has a non-normal distribution

Interpretation of Data:

From R11, mean, median and mode have approximately the same value that corresponds to 'De vez en cuando'. The measures of central tendency corresponds with the 50% percentile. The shape of the data shows that most of the values have approximately the same percentage of occurrence and concentrate to the 28.78%.

3.1.17 R12 : ¿Con qué frecuencia has sentido que las dificultades se acumulan tanto que no puedes superarlas?

	Nominal Value	Categorical Value
0	Nunca	0
1	Casi nunca	1
2	De vez en cuando	2
3	A menudo	3
4	Muy a menudo	4



Mean: 1.8182

Median: 1.5

Mode: 1

Analysis:

Kurtosis: -0.4684, Platykurtic

Skewness: 0.7447, Right Assymetry (+)

Normality Test:

Saphiro-Wilk Test: $W = 0.8291$ $p\text{-value} = 1.3357795069168787e-05$

H1: Variable has a non-normal distribution

Interpretation of Data:

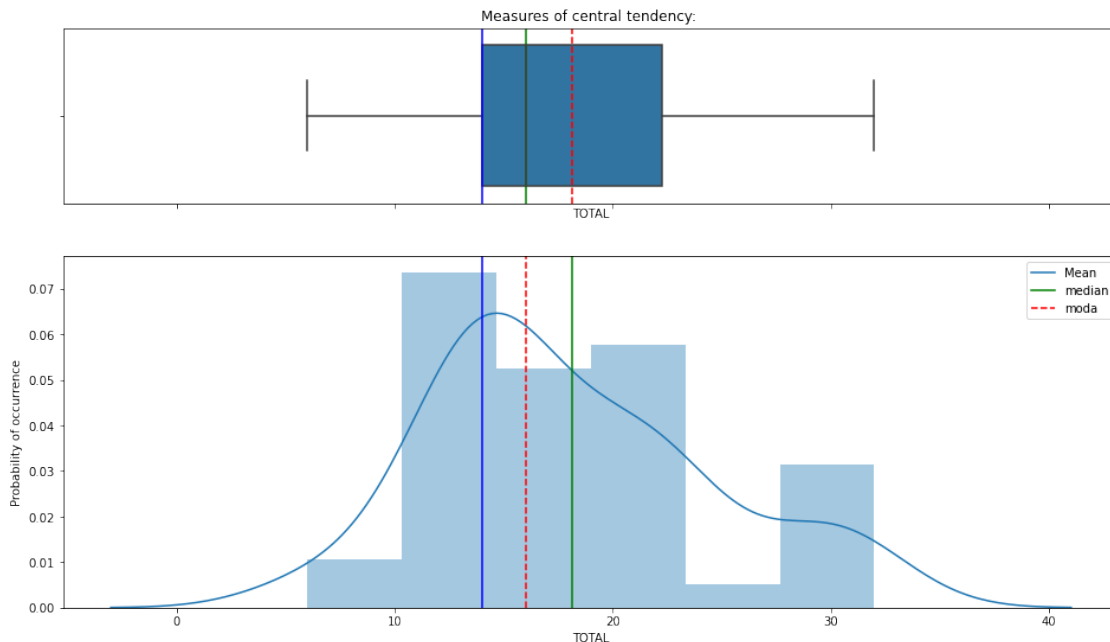
From R12, mean, median and mode have different values, mode corresponds to 'Casi nunca' value - 47.7% of occurrence percentage and median and mode are more near to the 'De vez en cuando' value - 25% of occurrence percentage. Distribution is skewed to 'Casi nunca' value that is approximately half of the data.

3.2 Dependent Variable

```
[9]: # Calculate measures of central tendency for the Dependent Variable
```

```
ds.measures_of_center(df, dependent_variable)
```

3.2.1 TOTAL



Mean: 18.1136

Median: 16.0

Mode: 14

Analysis:

Kurtosis: -0.3722, Platykurtic

Skewness: 0.545, Right Assymetry (+)

Normality Test:

Saphiro-Wilk Test: $W = 0.9386$ $p\text{-value} = 0.021099934354424477$

H1: Variable has a non-normal distribution

Interpretation of Data:

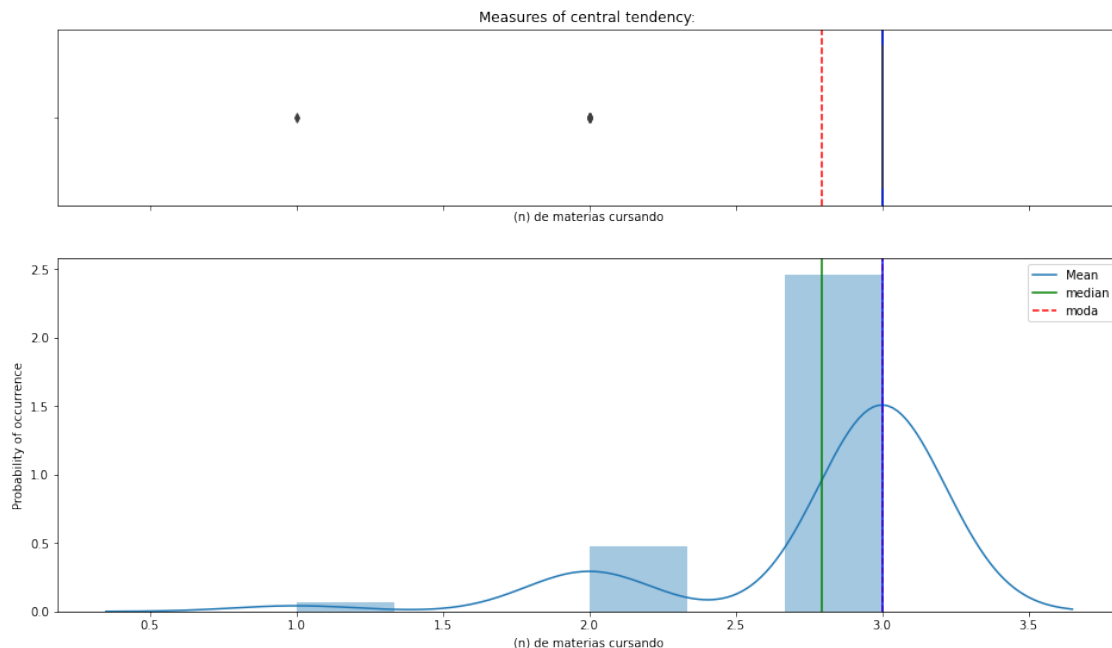
From total, mean, median and mode have different values, however mean concentrate in the 18 score, lowest score was 6 and highest 32, but most the data in the 10 to 25 score range, that means that comparing with the lowest and highest value most of data is on the center and the assumption is that the perceived stress is in a normal parameter.

3.3 Intervening Variable

```
[10]: # Calculate measures of central tendency for the Dependent Variable
```

```
ds.measures_of_center(df, intervening_variable)
```

3.3.1 (n) de materias cursando



Mean: 2.7955

Median: 3.0

Mode: 3

Analysis:

Kurtosis: 3.9536, Leptokurtic

Skewness: -2.1545, Left Assymetry (+)

Normality Test:

Saphiro-Wilk Test: W = 0.4906 p-value = 3.735199302634662e-11

H1: Variable has a non-normal distribution

Interpretation of Data:

From (n) of coursing subjects, the mean, median and mode have approximately the same value. Most of the data concentrate in the third value, that means, most of the subjects of the sample were coursing 3 subjects

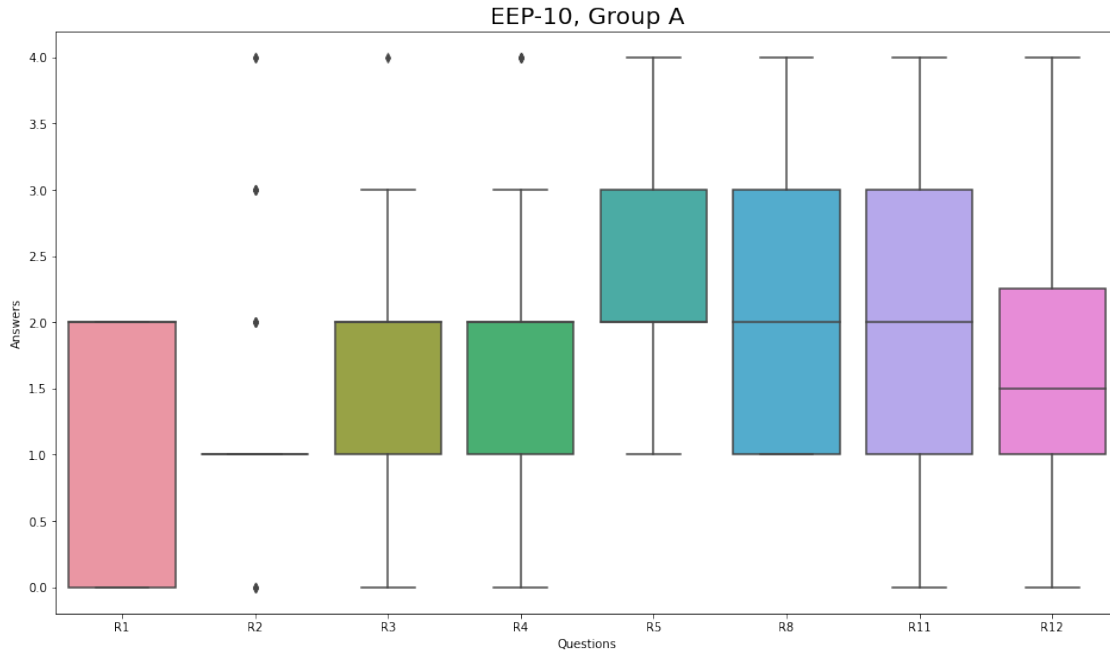
3.4 Boxplot for EEP-10 Instrument Responses

3.4.1 EEP-10 Responses, Group A: R1, R2, R3, R4, R5, R8, R11, R12

```
[11]: ds.boxplot_instrument_responses(df, "A")
```

	Key	Value
0	R1	¿Dónde percibes con mayor frecuencia estrés?
1	R2	En el ambiente universitario ¿Qué aspecto te provoca más estrés?
2	R3	Respecto a las clases de la Licenciatura en Psicología ¿Con qué frecuencia has estado afectado por alg
3	R4	¿Con qué frecuencia te has sentido incapaz de controlar las cosas importantes en tu vida?
4	R5	¿Con qué frecuencia te has sentido nervioso o estresado?
5	R8	¿Con qué frecuencia has sentido que no podías afrontar todas las cosas que tenías que hacer?
6	R11	¿Con qué frecuencia has estado enfadado porque las cosas que te han ocurrido estaban fuera de tu co
7	R12	¿Con qué frecuencia has sentido que las dificultades se acumulan tanto que no puedes superarlas?

	Nominal Value	Categorical Value
0	Nunca	0
1	Casi nunca	1
2	De vez en cuando	2
3	A menudo	3
4	Muy a menudo	4

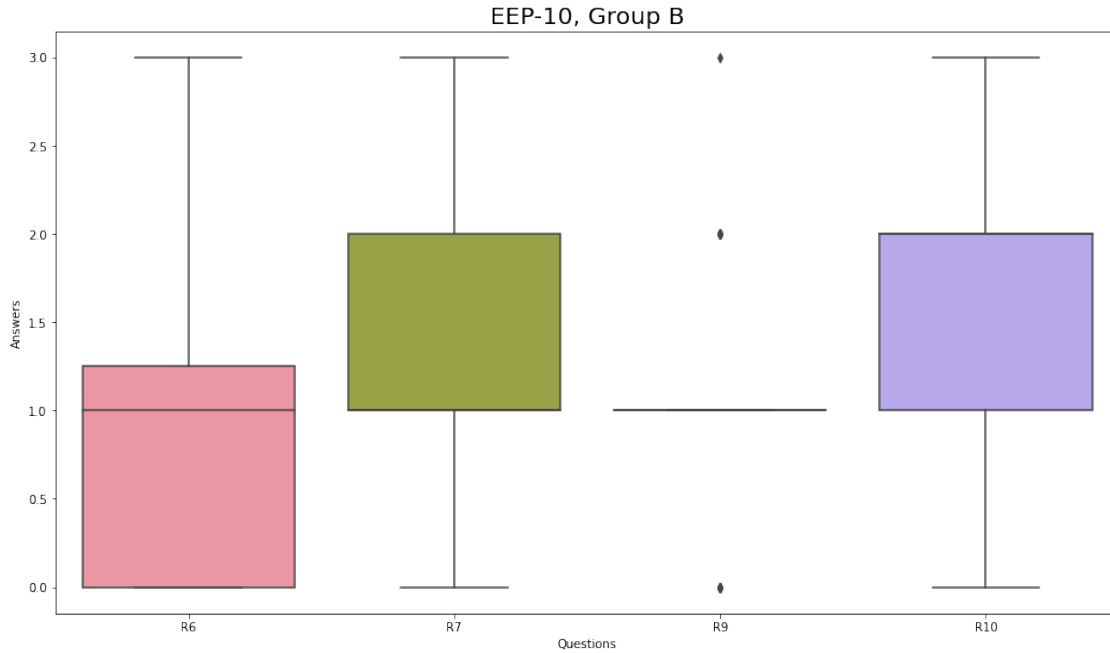


3.4.2 EEP-10 Responses, Group B: R6, R7, R9, R10, R11

[12]: `ds.boxplot_instrument_responses(df, "B")`

	Key	Value
0	R6	¿Con qué frecuencia ha estado seguro sobre su capacidad para manejar sus problemas personales?
1	R7	¿Con qué frecuencia has sentido que las cosas te van bien?
2	R9	¿Con qué frecuencia has podido controlar las dificultades en tu vida?
3	R10	¿Con qué frecuencia te has sentido que tenía todo bajo control?

	Nominal Value	Categorical Value
0	Nunca	4
1	Casi nunca	3
2	De vez en cuando	2
3	A menudo	1
4	Muy a menudo	0

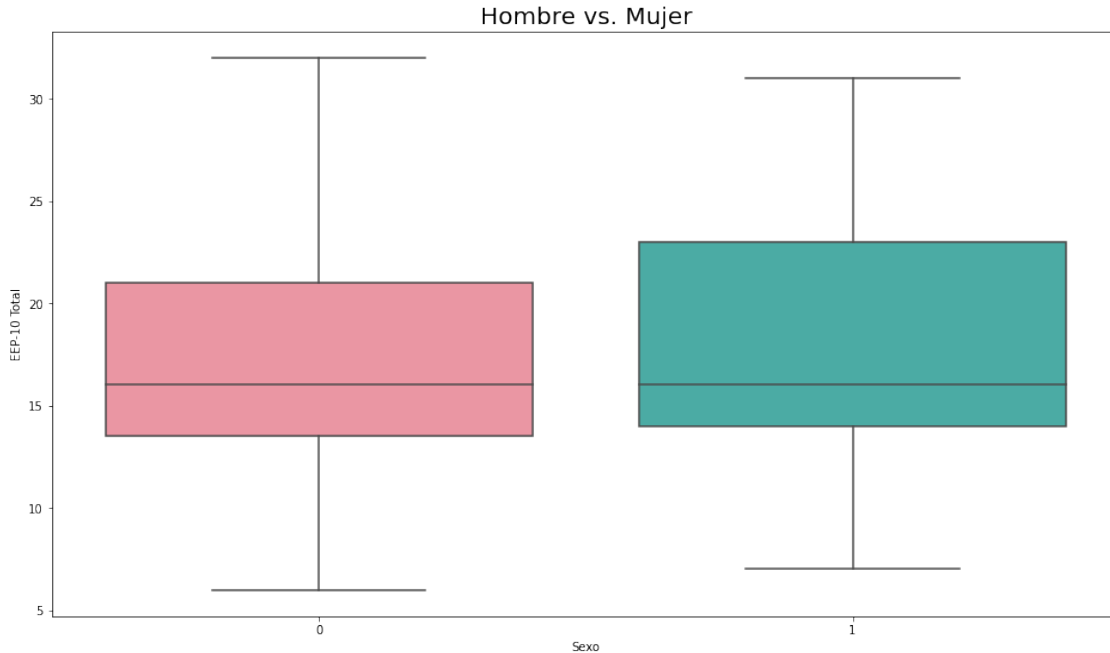


The data of the boxplot responses groups show several outliers and a difference in the measures of central tendency (mean, median, mode).

3.4.3 EEP-10 Responses, Variable 'Sexo' comparison

```
[13]: ds.data_variable_boxplot(df, 'Sexo', 'TOTAL', 'Hombre vs. Mujer', 'Sexo',
    ↪ 'EEP-10 Total')
```

	Nominal Value	Categorical Value
0	Hombre	0
1	Mujer	1



Comparing 'Sexo' variable, mean for 'Mujeres' and 'Hombres' has the same value, also the shape of the box is very similar with the difference that in the sample there were more women (75%) than men (25%).

4 Measures of dispersion

4.1 Independent Variables

```
[14]: # Calculate measures of dispersion for the Independent Variables
```

```
for independent_variable in independent_variables:
    ds.measures_of_dispersion(df, independent_variable)
```

```
Sexo:
Standard Deviation: 0.4380187953915666
Variance: 0.19186046511627908
```

```
Edad:
Standard Deviation: 9.097893722740162
Variance: 82.77167019027483
```

```
Nivel de estudios:
Standard Deviation: 1.132594170930334
Variance: 1.2827695560253705
```

Escolaridad:
Standard Deviation: 3.2961634434731533
Variance: 10.864693446088795

Cuatrimestre Cursando:
Standard Deviation: 1.119451315313576
Variance: 1.2531712473572951

R1:
Standard Deviation: 0.9386576714124694
Variance: 0.8810782241014793

R2:
Standard Deviation: 0.9135461653329312
Variance: 0.8345665961945032

R3:
Standard Deviation: 0.8611291078465235
Variance: 0.7415433403805496

R4:
Standard Deviation: 1.0198868237174104
Variance: 1.040169133192388

R5:
Standard Deviation: 0.8995535655735866
Variance: 0.8091966173361529

R6:
Standard Deviation: 0.7920722413504784
Variance: 0.6273784355179705

R7:
Standard Deviation: 0.8515624242343391
Variance: 0.7251585623678646

R8:
Standard Deviation: 0.9786300127743622
Variance: 0.9577167019027484

R9:
Standard Deviation: 0.6611381051492481
Variance: 0.4371035940803382

R10:
Standard Deviation: 0.7499119046217668
Variance: 0.5623678646934459

R11:
Standard Deviation: 1.0555401010206003
Variance: 1.1141649048625788

R12:
Standard Deviation: 1.0404097420197684
Variance: 1.082452431289641

4.2 Dependent Variable

```
[15]: # Calculate measures of dispersion for the Dependent Variable  
  
ds.measures_of_dispersion(df, dependent_variable)
```

TOTAL:
Standard Deviation: 6.412980691773955
Variance: 41.12632135306555

4.3 Intervening Variable

```
[16]: # Calculate measures of dispersion for the Dependent Variable  
  
ds.measures_of_dispersion(df, dependent_variable)
```

TOTAL:
Standard Deviation: 6.412980691773955
Variance: 41.12632135306555

5 Conclusion

There is a high dispersion within the variables due to the variability found in some of the data, to continue with the correlational variable analysis and therefore with the inferential statistics, non-parametric data analysis should be performed.

An alternative we recommend is to apply again the measuring instrument and select a more homogenous sample or remove the outliers to try to normalize the non-normal distributions and continue with the parametric analysis, although, have more control with the variables that we are measuring.