* we need to sort our data before the start calculations

1-Explain dataset and explain the column which you choose and why you choose

I choosed Top50 Spotify Song –Liveness column, because i use spotify and i love listen to musiz so i wonder abut songs liveness.

2- Find mean of column data

Mean is measuring the average value of a sample, we can explain like this; 5.0, 6.0, 6.0, 6.0, 6.0, 7.0, 7.0, 7.0, 7.0, 7.0, 8.0, 8.0, 8.0, 8.0, 8.0, 9.0, 9.0, 9.0, 9.0, 10.0, 10.0, 10.0, 11.0, 11.0, 11.0, 11.0, 11.0, 11.0, 12.0, 12.0, 12.0, 12.0, 13.0, 13.0, 14.0, 14.0, 15.0, 16.0, 16.0, 17.0, 18.0, 19.0, 23.0, 24.0, 32.0, 36.0, 36.0, 41.0, 44.0, 58.0 these are our values, and there are 50 values, so n=50

$(5.0+6.0+6.0+6.0.........+44.0+58.0) / n$ ---> $733.0/50 = 14.66$ Mean

3- Find median of column data

Median is measuring the central value, when we find median we must pay attenion to size if n is even in that case we find the middle pair of numbers, and then find the value that is half way between them it means adding them together and dividing by two. If n is odd place the numbers in value order and find the middle. In our vules we have 50 values, so we will find the middle pair of numbers these are 11.0 and 11.0

$(11.0+11.0)/2 = 11.0$ ---> Median

4- Find the variance, standard deviation and standard error

These values are my population, so i will use population variance formula; $\sigma^2 = \Sigma(x_i - \mu)^2 / n$

We know the mean is $=14.66$, n=50

$[(5.0-14.66)^2 + (6.0-14.66)^2 + (6.0-14.66)^2 + \ldots\ldots + (44.0-14.66)^2 + (58.0-14.66)^2] / 50 = 121.14444$ Variance

Poupulation standart devision is σ, so $\sqrt{121.1444} = 11.006$

Population Standart error formula ; $\sigma /(\sqrt{n})$ ---> $11.006/\sqrt{50} = 1.556$

5- Decide the shape of distribution

If mean and median equal --> symmetric

If Mean < Median --> left-skewed

If Mean > Median --> right -skewed

Our Mean is = 14.66 Median is =11.0   Mean > Median, so its rigt-skewed also it means there are less data than average. We can think like this; if this would be result of exam, students do not get successful results and we can say maybe this was a hard exam.

6- Find outliers if there is

If we want to find outliers firstly we have to find quarties (q1, and q3), also we know the q2 equls to median

When find the quarties, we should chechk n is odd or even?

If n is even, for the q2=meadian i explain how can we find 2nd question, for the q1 and q3 i will explain with small size for to be clearly

For example 4,5,7,8,12,14,14,15,16,20 this is my values q2=12+14/2=13 for to find q1 and q3 we can seperate between 12 and 14 like this 4,5,7,8,12 --- 14,14,15,16,20 and now we can find like how we find median, in left side in the middle is 7, so q1=7, For the right side in the middle is 15, so q3=15

Also we need to find iqr, iqr = q3-q1 = 15-7=8

Now we can find our interval [q1 - (1.5 x iqr) , q3 + (1.5 x iqr)]

7-(1.5 x 8) = -5

15+(1.5 x 8)= 27

For this example our interval is [-5 , 27] and when we check our values there i no any outliers

If n is odd, for the q2=median we will do it same thing how can find median, for the q1 and q3 i will explain again example; 4,5,7,8,12,14,14,15,16 these are my values in the middle 12 is my median and also q2, for the find q1 and q3 we can seperate from 12 but 12 not include like this; 4,5.7,8 ---- 12---- 14,14,15,16 and we can find from left sight q1 like median (7+5)/2 =6 , from the right side (14+15)/2 = 14.5

Iqr=q3-q1 = 14.5 - 6=8.5

our interval [q1 - (1.5 x iqr) , q3 + (1.5 x iqr)]

6-(1.5 x 8.5) = -6.75

14.5 +(1.5 x 8.5) = 27.25

For this example our interval is [-6.75 , 27.25] and when we check our values there i no any outliers

When back our population n is even, in this case we calculated median and we know q2 equls to median, so median=q2=11.0

When follow the steps we will find q1=8.0 and q3=16.0   Iqr=16.0-8.0 = 8.0

our interval [q1 - (1.5 x iqr) , q3 + (1.5 x iqr)]

8.0 - (1.5 x 8.0) = 4.0

16.0 +(1.5 x 8.0) = 28.0

For my populations interval is [4.0, 28.0] and when we check my values 32.0 , 36.0 ,41.0 , 44.0 ,58.0 numbers are outliers

7- Graph the column data using histogram and make comment about data

Histogram is; shows the shape of a pmf or a pdf of data, checks for homogeneity, and suggests possible outliers. To construct a histogram, we split the range of data into equal intervals, "bins," and count how many observations fall into each bin.

When we want to draw fisrtly we should find number of bin, bin is our number of column. To find the bin there some formula i use it $\log_2 n + 1$, we know the n=50

In these case $\log_2 50 + 1$ ---> 5,64 + 1 = 6,64 we can choose bin=7

And we will start our min value this is 5.0

32 observation between 5.0 and 12.0

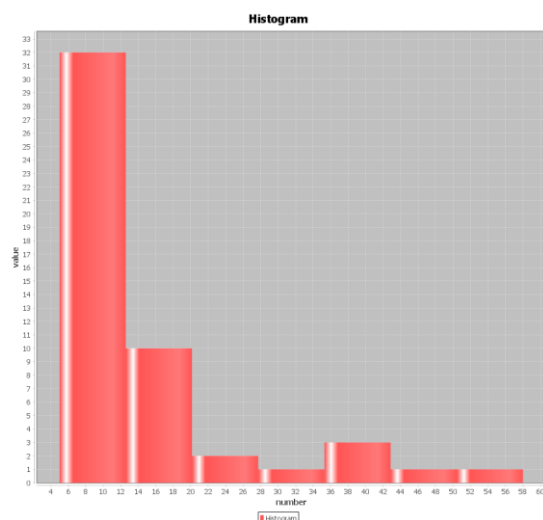10 observation between 13.0 and 20.0

2 observation between 21.0 and 28.0

1 observation between 29.0 and 35.0

3 observation between 36.0 and 43.0

1 observation between 44.0 and 51.0

1 observation between 52.0 and 58.0

8- Draw boxplot and make comment

Boxplot is; The main descriptive statistics of a sample can be represented graphically by a boxplot. To construct a boxplot, we draw a box between the first and the third quartiles, a line inside a box for a median, and extend whiskers to the smallest and the largest observations,

thus representing a so-called five-point summary: five point summary=(minXi,Q^1,M^ ,Q^3,maxXi)

We found our limits [-4.0 , 28.0] in these limits our min valuse minXi=5.0 , mazXi=24.0, also we found too q1,q2 and q3, we know the q2=Median, and we know the shape of distribution is rigth-skewed
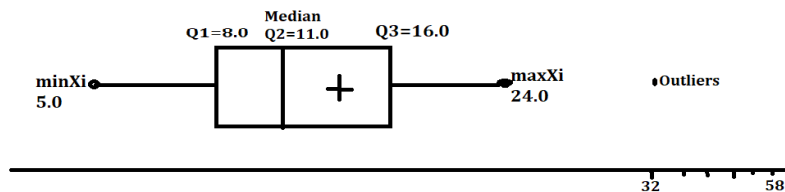
9- Take specific number of sample and construct %95 confidence interval for the mean and variance

For confidence interval i select these numbers 19.0,18.0,15.0,11.0, 9.0, 6.0

Fisrstly i will solve confidence interval for mean, and confidence interval for mean formula; $X' \pm Z_{\alpha/2} . \sigma / \sqrt{n}$ i will use this formula because i know my population standard deviation σ =11.006

And i use a population standard deviation anf my population size n=50, if n>30 we must use Ztest

For %95 1-α =%95, so α=0.05 we need α/2=0.025 --> when we chech from Ztable $Z_{\alpha/2}$=1.96

Finally we need to find mean of our sample X'=(19.0+18.0+15.0+11.0+9.0+6.0)/6 = 13.0

Now we can use our formula 13.0 ± 1.96 x (11.006 / $\sqrt{6}$) --> 13.0 ± 8.806 --> [4 , 21] is our %95 confidence interval for mean.

%95 confidence interval for variance formula; $[(n-1)s^2 / X^2_{\alpha/2} \quad (n-1)s^2 / X^2_{1-\alpha/2}]$

Firslty need to find our sample variance s^2 -->[$(19.0 - 13.0)^2 + (18.0 - 13.0)^2 + (15.0 - 11.0)^2 + (11.0 - 13.0)^2 + (9.0 - 13.0)^2 + (6.0 - 13.0)^2$] / 5 we divide 5 because when we calculate sample variance ve must divide n-1, so s^2 =26.8

 For %95 1-α =%95, so α=0.05 we need α/2=0.025 and 1- α/2 =0.975

$X^2_{\alpha/2} = X^2_{0.025}$ = 12.8, $X^2_{1-\alpha/2} = X^2_{0.975}$ =0.83 --> we find from Chi-Square Table, when we chech table we need to knor df, df is = n-1, so our df is =5

Know we know enough information about our formula and we can apply,

--> [(5x 26.8)/12.8 , (5x26.8) /0.83] --> [10 , 161] is our %95 confidence interval for variance

10- How large a sample for your data should be collected to estimate the population mean with a margin at most 0.1 units with confidence 90%.

For this choice i will keep on from 9th question, and i use formula $= n \geq [(Z_{\alpha/2} \times \sigma) / \Delta]^2$

We know the $\sigma = 11.006$

$\Delta = 0.1$

For %90 1-α =%90 , α=0.1   , $\alpha/2$=0.05   , $Z_{0.05} = 1.645$ from Z table

We can use formula $n \geq [(1.645 \times 11.006) / 0.1]^2$, $n \geq 32,781$, so n should be n=32,782