**Bedroom**

This dependet factor tell us how many bedrooms there are in a house.

According to the basic statistic the rage in from 0 bedrooms to a maximum of 33.

In the table is shown the repartition of the data.

|  |  |
| --- | --- |
| **number of bedrooms** | **number of houses** |
| 0 | 13 |
| 1 | 199 |
| 2 | 2760 |
| 3 | 9824 |
| 4 | 6882 |
| 5 | 1601 |
| 6 | 272 |
| 7 | 38 |
| 8 | 13 |
| 9 | 6 |
| 10 | 3 |
| 11 | 1 |
| 33 | 1 |

**Basic Statistic**

In the table below there is the statistic for the price according to the number of bedrooms of a house.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | 0 | 1 | 2 | 3 |
| Mean | 4,10E+05 | 3,18E+05 | 4,01E+05 | 4,66E+05 |
| Median | 2,88E+05 | 2,99E+05 | 3,74E+05 | 4,13E+05 |
| Variance | 1,29E+11 | 2,22E+10 | 3,92E+10 | 6,89E+10 |
| Std. Deviation | 3,59E+05 | 1,49E+05 | 1,98E+05 | 2,62E+05 |
| Minimum | 1,40E+05 | 7,50E+04 | 7,80E+04 | 8,20E+04 |
| Maximum | 1,30E+06 | 1,25E+06 | 3,28E+06 | 3,80E+06 |
| Range | 1,16E+06 | 1,17E+06 | 3,20E+06 | 3,72E+06 |

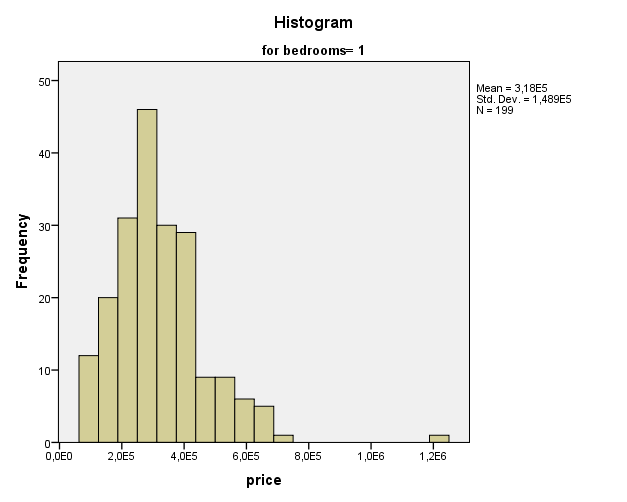
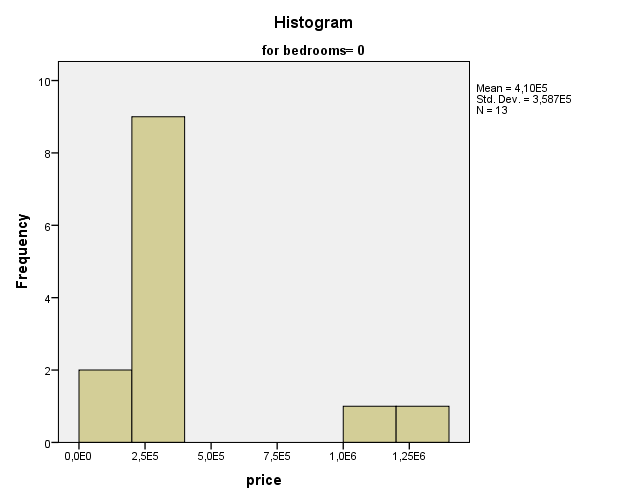
|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | 4 | 5 | 6 | 7 |
| Mean | 6,35E+05 | 7,87E+05 | 8,26E+05 | 9,51E+05 |
| Median | 5,50E+05 | 6,20E+05 | 6,50E+05 | 7,29E+05 |
| Variance | 1,51E+11 | 3,55E+11 | 6,39E+11 | 5,48E+11 |
| Std. Deviation | 3,89E+05 | 5,96E+05 | 7,99E+05 | 7,40E+05 |
| Minimum | 1,00E+05 | 1,33E+05 | 1,75E+05 | 2,80E+05 |
| Maximum | 4,49E+06 | 7,06E+06 | 7,70E+06 | 3,20E+06 |
| Range | 4,39E+06 | 6,93E+06 | 7,53E+06 | 2,92E+06 |

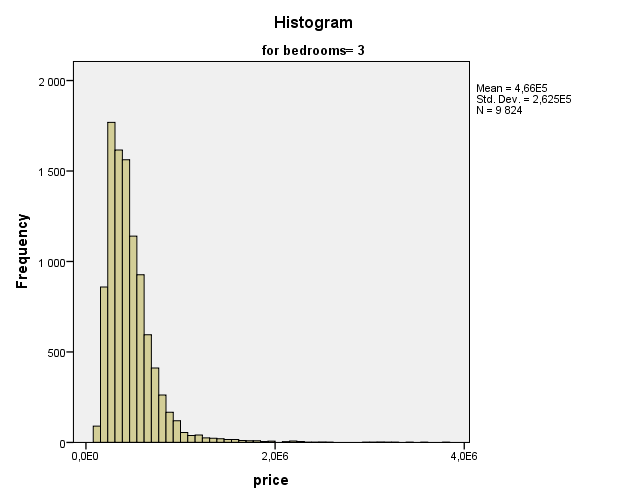
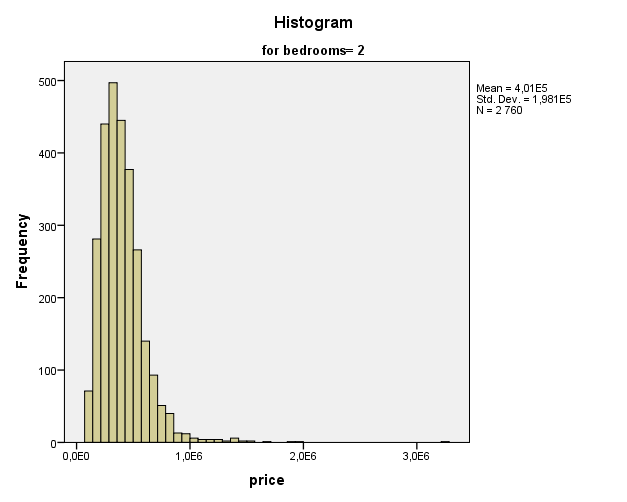
|  |  |  |  |
| --- | --- | --- | --- |
|  | 8 | 9 | 10 |
| Mean | 1,11E+06 | 8,94E+05 | 8,19E+05 |
| Median | 7,00E+05 | 8,17E+05 | 6,60E+05 |
| Variance | 8,05E+11 | 1,46E+11 | 8,10E+10 |
| Std. Deviation | 8,97E+05 | 3,82E+05 | 2,85E+05 |
| Minimum | 3,40E+05 | 4,50E+05 | 6,50E+05 |
| Maximum | 3,30E+06 | 1,40E+06 | 1,15E+06 |
| Range | 2,96E+06 | 9,50E+05 | 4,98E+05 |

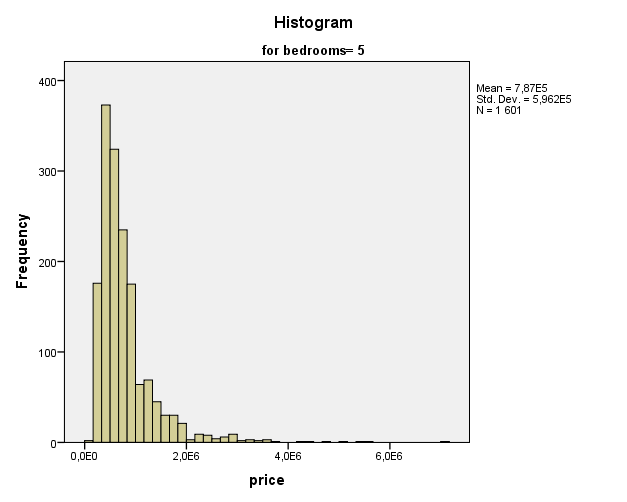
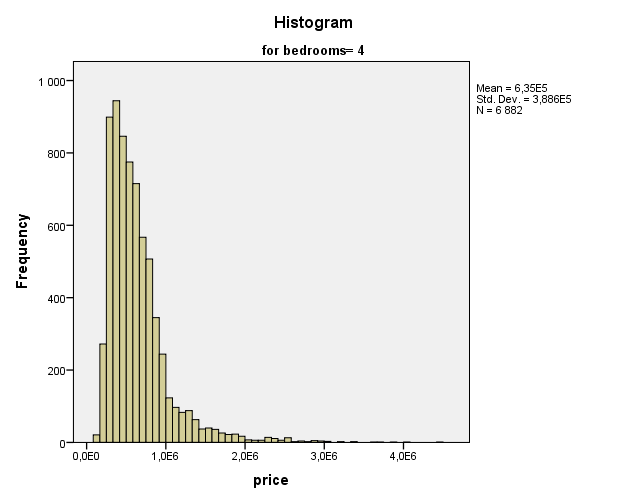
For what concern the house with 11 and 33 bedrooms we don’t have any statistic because the number of house is 1 and so the price is constant.

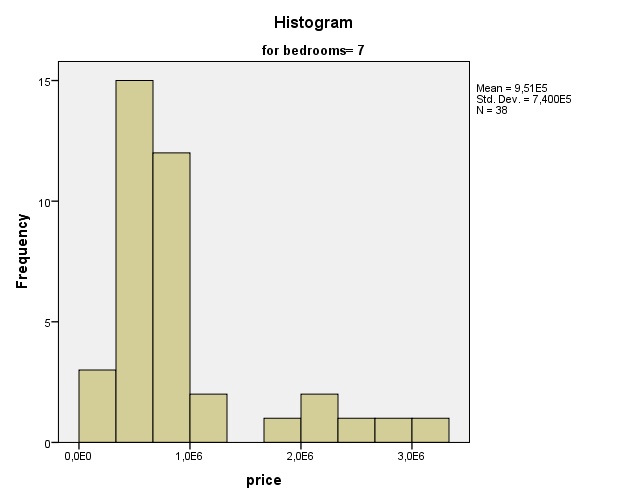
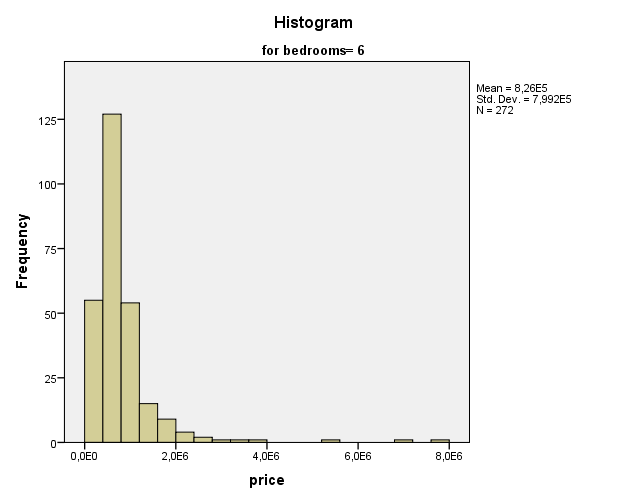
We can say looking at the mean that the price of the houses rise with the number of rooms. There is an exception for the houses with 8 rooms, but as shown in the table the standard deviation for this case is higher than in the others.

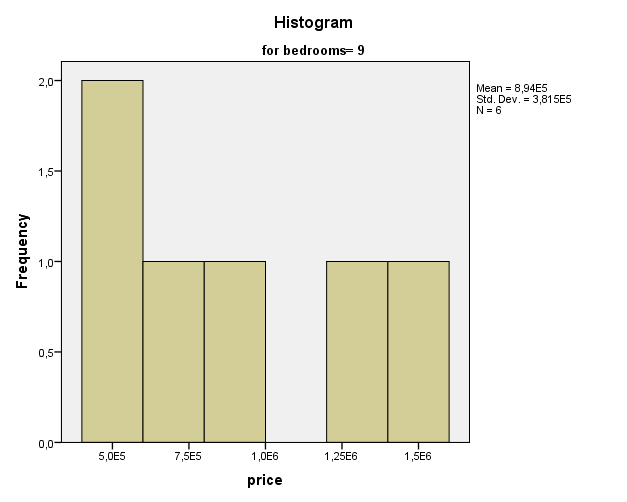
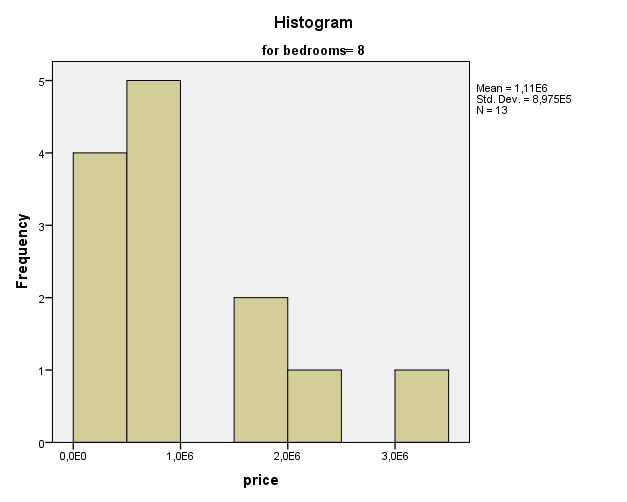
**Distribution of prices**

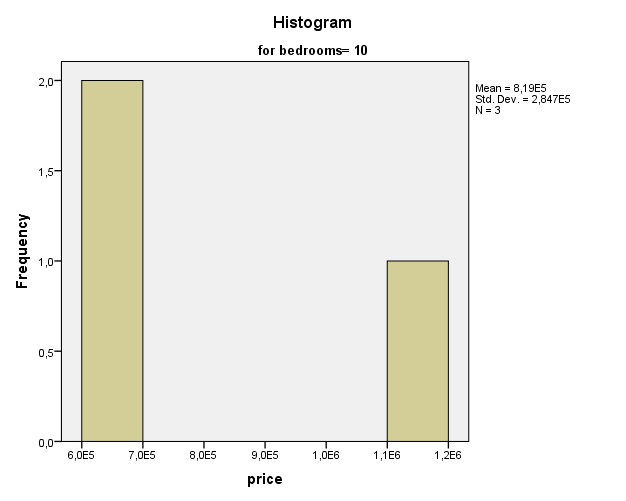








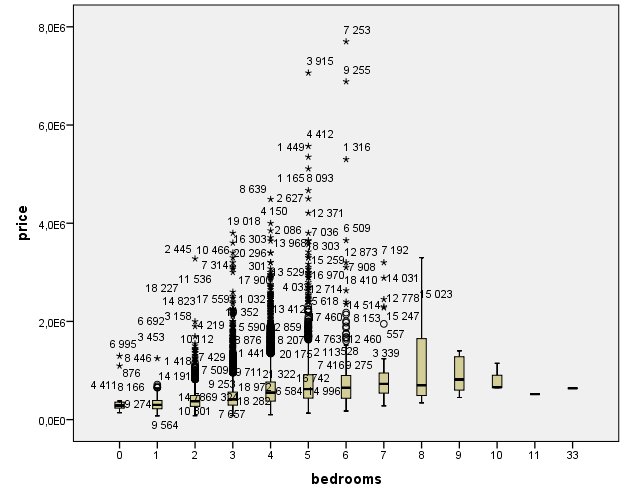




The distribution for the price of the house from 1 to 6 rooms are similar. For what concern the others, because the number of the houses is so small we have different types of distribution. For the other categories of house, the pattern seems to be the same, but the number of data in lower, so the distributions has some missing point. However, we can say for each category that as the price increase the number of houses decrease.

We don’t have any histogram for bedroom= 11, 33 because we the price is constant (only one house for each category)

**Outliers**



We don’t have any below outliers probably because an house can’t be sold for less than 0 $.

The highest number of outliers is concentrated between the houses with 2 and 7 rooms. But we can say that the percentage of outliers is low.

Below we have the 5 highest outliers for house price by bedrooms.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  |  | 1 | 2 | 3 | 4 | 5 |
| 0 | ID | 6995 | 876 | 3120 | 9774 | 8478 |
| Price | 1,30E+06 | 1,10E+06 | 3,80E+05 | 3,55E+05 | 3,40E+05 |
| 1 | ID | 8446 | 9889 | 4411 | 8166 | 19274 |
| Price | 1,25E+06 | 7,13E+05 | 6,80E+05 | 6,55E+05 | 6,55E+05 |
| 2 | ID | 2445 | 6692 | 3158 | 18076 | 13420 |
| Price | 3,28E+06 | 2,00E+06 | 1,90E+06 | 1,70E+06 | 1,55E+06 |
| 3 | ID | 19018 | 2865 | 7990 | 18200 | 11536 |
| Price | 3,80E+06 | 3,60E+06 | 3,40E+06 | 3,30E+06 | 3,20E+06 |
| 4 | ID | 8639 | 4150 | 2086 | 16303 | 15256 |
| Price | 4,49E+06 | 4,00E+06 | 3,85E+06 | 3,71E+06 | 3,64E+06 |
| 5 | ID | 3915 | 4412 | 1449 | 1165 | 8093 |
| Price | 7,06E+06 | 5,57E+06 | 5,35E+06 | 5,11E+06 | 4,67E+06 |
| 6 | ID | 7253 | 9255 | 1316 | 6509 | 7908 |
| Price | 7,70E+06 | 6,89E+06 | 5,30E+06 | 3,65E+06 | 3,20E+06 |
| 7 | ID | 7192 | 14031 | 15023 | 15247 | 12778 |
| Price | 3,20E+06 | 2,89E+06 | 2,45E+06 | 2,30E+06 | 2,28E+06 |
| 8 | ID | 18478 | 4036 | 17236 | 10959 | 9453 |
| Price | 3,30E+06 | 2,15E+06 | 1,97E+06 | 1,65E+06 | 9,00E+05 |
| 9 | ID | 16845 | 6080 | 18444 |  |  |
| Price | 1,40E+06 | 1,28E+06 | 9,34E+05 |  |  |
| 10 | ID | 13315 |  |  |  |  |
| Price | 1,15E+06 |  |  |  |  |

**Floor**

This factor tells us how many floors has each house. According to the basic statistic, the minimum value is 0 and the maximum value is 3,5 (the half part of the number indicate that an house has an attic).

In the table is shown the repartition of the data.

|  |  |
| --- | --- |
| Number of floors | Number of houses |
| 1 | 10680 |
| 1,5 | 1910 |
| 2 | 8241 |
| 2,5 | 161 |
| 3 | 613 |
| 3,5 | 8 |

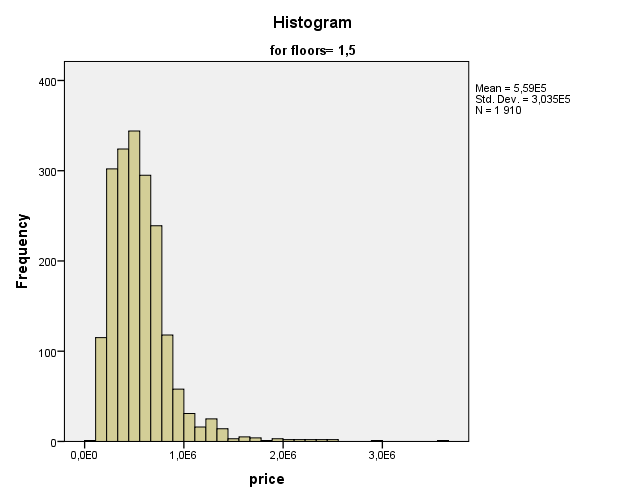
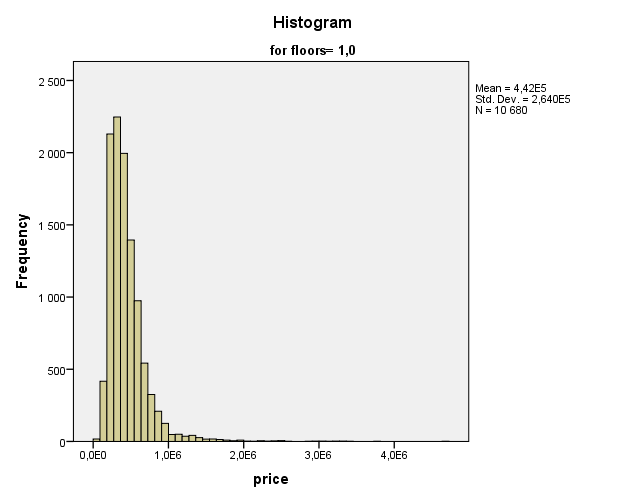
**Basic Statistic**

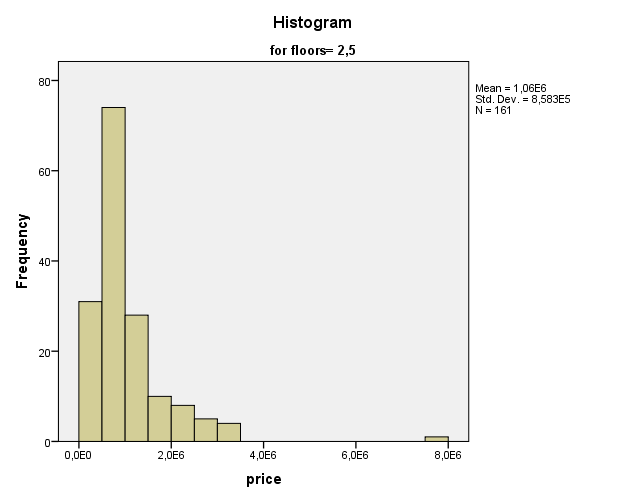
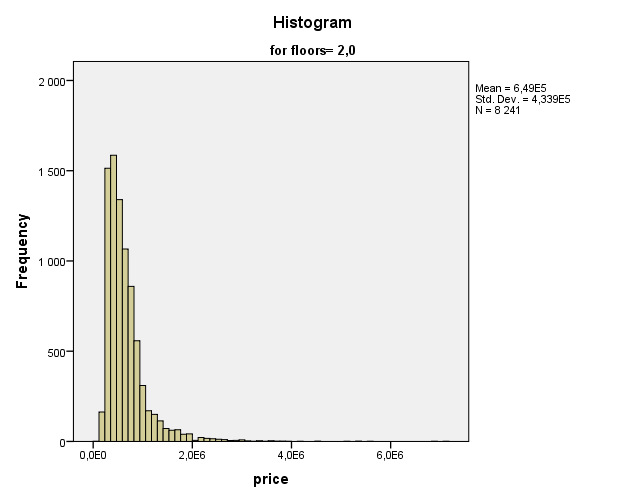
|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | 1 | 1,5 | 2 | 2,5 | 3 | 3,5 |
| Mean | 4,4E+05 | 5,6E+05 | 6,5E+05 | 1,1E+06 | 5,8E+05 | 9,3E+05 |
| Median | 3,9E+05 | 5,2E+05 | 5,4E+05 | 8,0E+05 | 4,9E+05 | 5,3E+05 |
| Variance | 7,0E+10 | 9,2E+10 | 1,9E+11 | 7,4E+11 | 1,1E+11 | 9,6E+11 |
| Std. Deviation | 2,6E+05 | 3,0E+05 | 4,3E+05 | 8,6E+05 | 3,4E+05 | 9,8E+05 |
| Minimum | 7,5E+04 | 9,2E+04 | 9,0E+04 | 2,6E+05 | 2,1E+05 | 4,4E+05 |
| Maximum | 4,7E+06 | 3,7E+06 | 7,1E+06 | 7,7E+06 | 3,1E+06 | 3,3E+06 |
| Range | 4,6E+06 | 3,6E+06 | 7,0E+06 | 7,4E+06 | 2,9E+06 | 2,9E+06 |

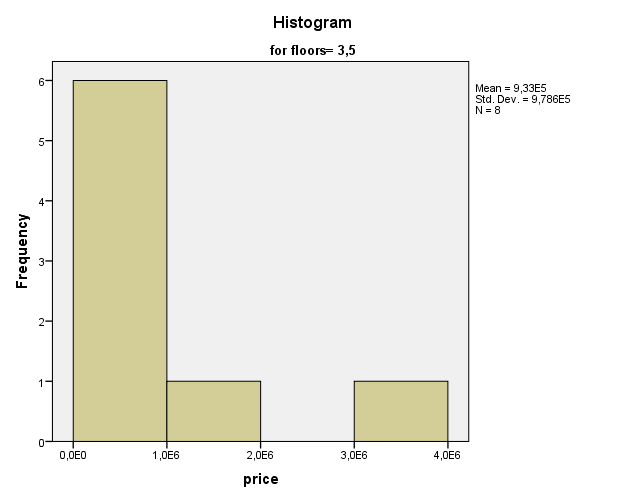
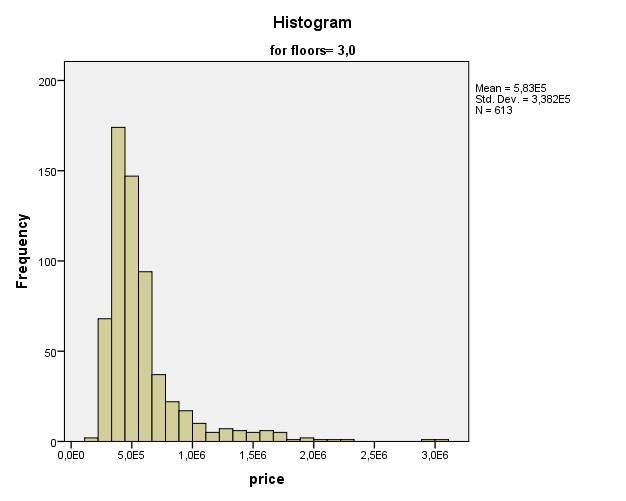
At first glance, as the number of floors increase also the prices increases. We have an exception for the houses with two floors and the attic, but if we look at the standard deviation, that is also higher than the others.

Also, the houses with three floors seems to be different, in fact we have a mean price lower than the previous ones.

**Distribution of prices**

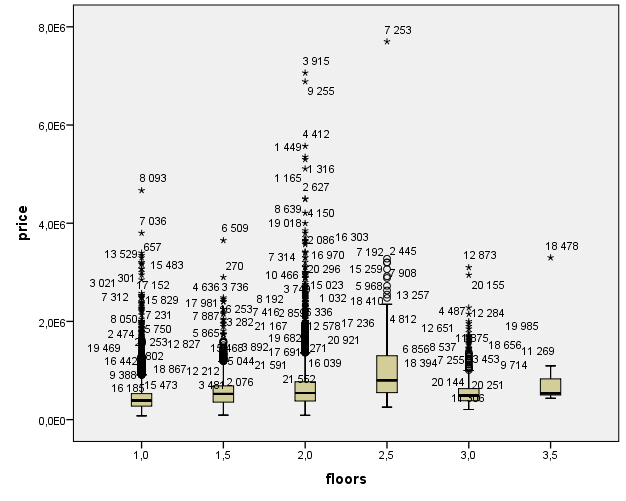






All the distributions are similar, as said before the only thing that change is the mean.

**Outliers**



From this graphic we can see that the higher number of outliers is within the house from 1 to 2 floors. This because in these 3 categories we have the higher number of houses. As we can see however, the number of outlier is not elevated (low percentage of that).

In the table below we will highlights the 5 highest outliers for each categories of house (based on the number of floors)

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  |  | 1 | 2 | 3 | 4 | 5 |
| 1 | ID | 8093 | 7036 | 13968 | 20461 | 18200 |
| Price | 4,67E+06 | 3,80E+06 | 3,40E+06 | 3,35E+06 | 3,30E+06 |
| 1,5 | ID | 6509 | 270 | 4636 | 3736 | 12714 |
| Price | 3,65E+06 | 2,90E+06 | 2,49E+06 | 2,45E+06 | 2,41E+06 |
| 2 | ID | 3915 | 9255 | 4412 | 1449 | 1316 |
| Price | 7,06E+06 | 6,89E+06 | 5,57E+06 | 5,35E+06 | 5,30E+06 |
| 2,5 | ID | 7253 | 2445 | 7192 | 7908 | 15259 |
| Price | 7,70E+06 | 3,28E+06 | 3,20E+06 | 3,20E+06 | 3,07E+06 |
| 3 | ID | 12873 | 20155 | 12778 | 7701 | 4487 |
| Price | 3,10E+06 | 2,95E+06 | 2,28E+06 | 2,15E+06 | 2,00E+06 |
| 3,5 | ID | 18478 | 876 | 20773 | 11595 |  |
| Price | 3,30E+06 | 1,10E+06 | 5,64E+05 | 5,44E+05 |  |

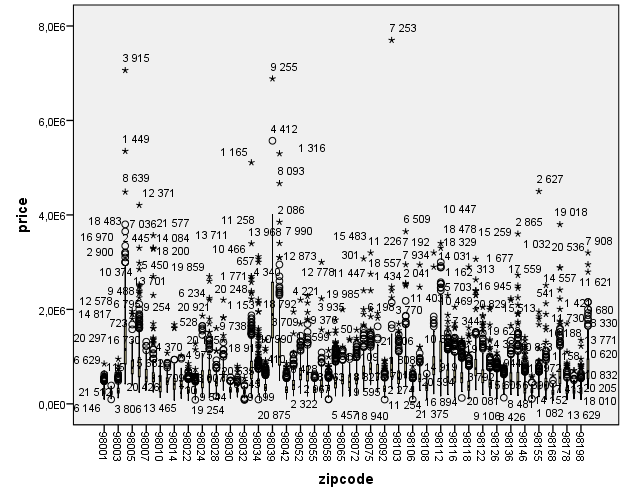
**ZIpCode**

The Zip code is a variable that indicates in which city the house is located. Because of that we consider this variable not as a number (how you call it?). Fot the King County the Zip code starts from 98001 and end at 98199. Because we have 199 classes, we will do a briefly analysis without presenting all the data.

The number of house is quite well distributed for each zip code. We had some exception of course; the minimum number of houses is 50 for the zip code = 98039 (city of Medina) and the maximum is 602 for the zip code 90103 (which correspond to the city of Seattle, the capital of the County).

The average price for the houses is 5,61E+05, the standard deviation is 3,00E+05. The maximum mean is for the zip code 90039(2,16E+06 $), the minimum mean is for the zip code 90002 (city of Auburn, 2,34E+05 $).

**Outliers**



From the analysis with boxplots we can see that not every zip codes have outliers. SPSS in fact shows us only those who have outliers. We can say that the outliers are well distributed with some values too much out the fences.

On average, we can say that every city present some outliers, so in every city we can find houses with prices outside the fence.