**YearBuilt**

This variable says when the house was built.

We have grouped the data in intervals. There are 7 interval : below 1900 ; 1900-1920 ;1920-1940 ;1940-1960 ;1960-1980 ;1980-2000 ;above 2000.

This is the repartition between each interval.

|  |  |
| --- | --- |
| Yr\_built | Number of house |
| <1900 | 87 |
| 1900-1920 | 1462 |
| 1920-1940 | 1780 |
| 1940-1960 | 4308 |
| 1960-1980 | 4937 |
| 1980-2000 | 4498 |
| >2000 | 4541 |

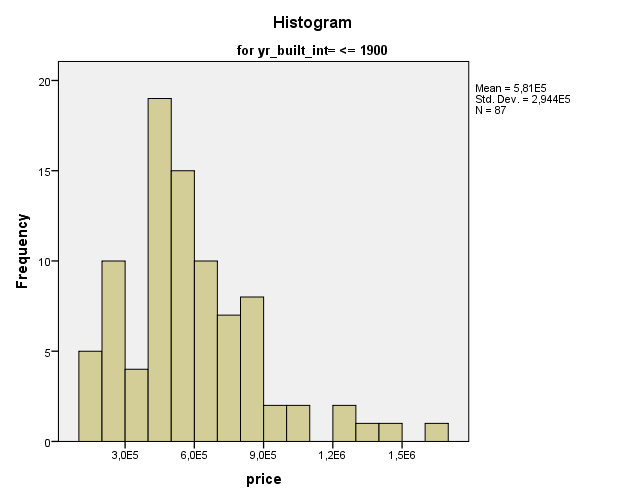
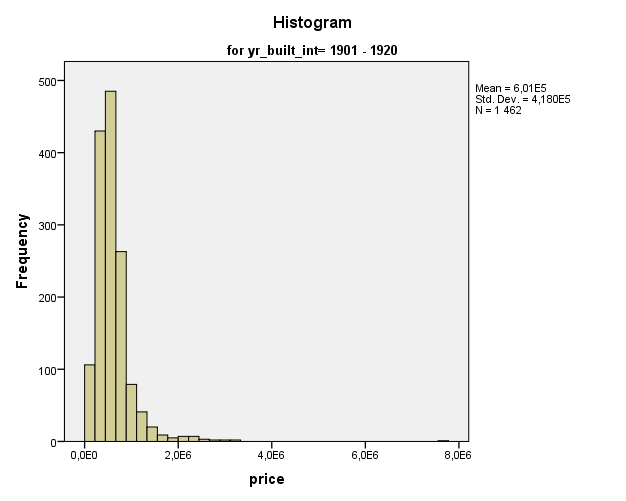
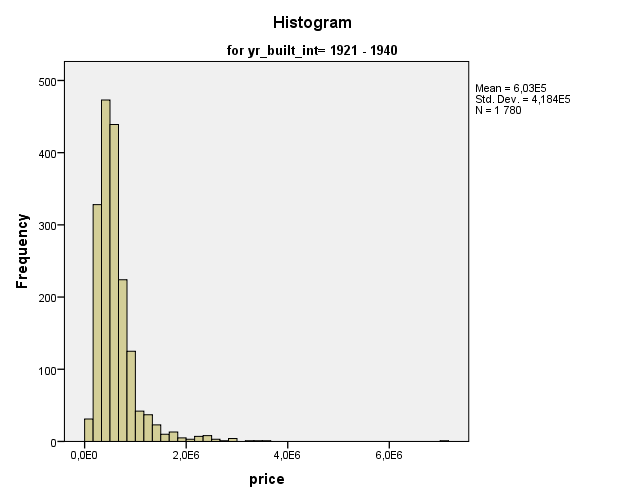
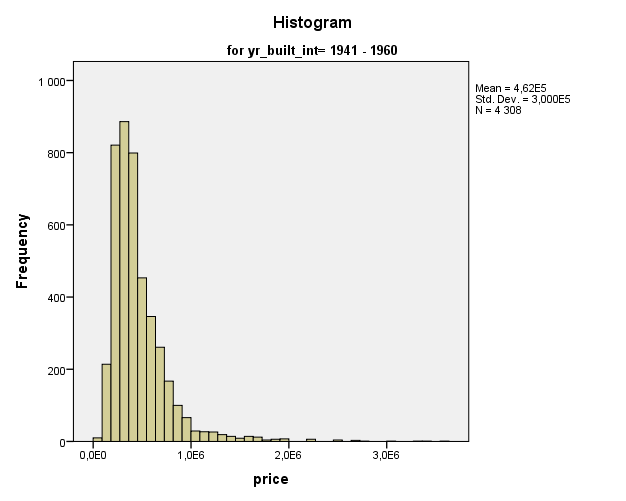
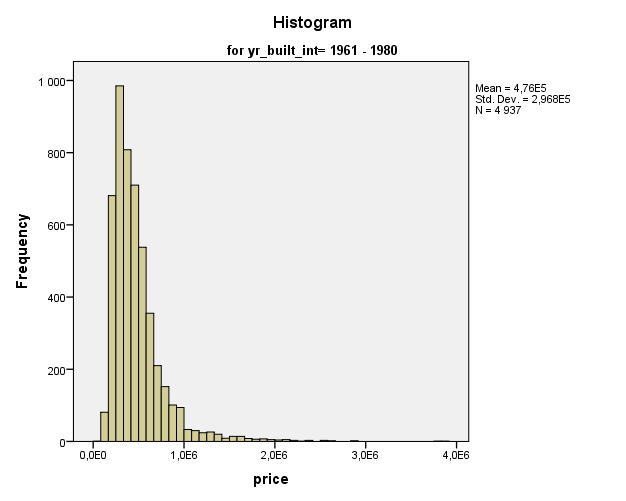
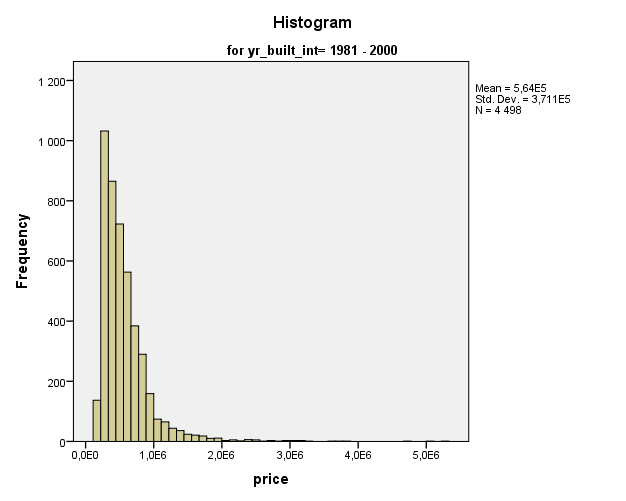
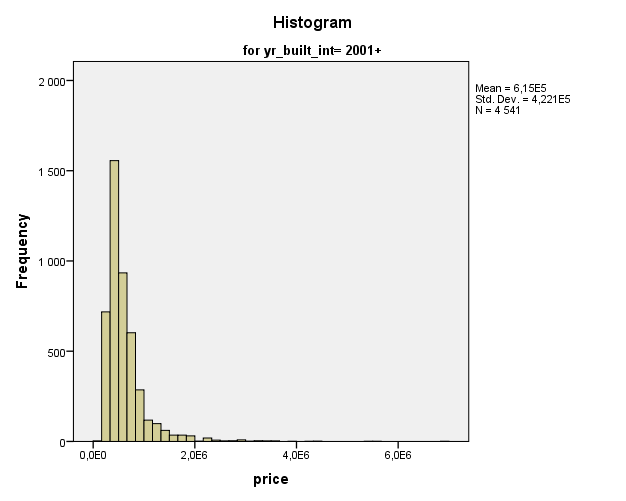
**Basic statistics**

There are the statistics for houses price ($) acording ofthe yr\_built:

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | <1900 | 1900-1920 | 1920-1940 | 1940-1960 | 1960-1980 | 1980-2000 | >2000 |
| Mean | 5,2\*10^5 | 6\*10^5 | 6\*10^5 | 4,6\*10^5 | 4,8\*10^5 | 5,6\*10^5 | 6,2\*10^5 |
| Median | 5,5\*10^5 | 5,3\*10^5 | 5,3\*10^5 | 3,9\*10^5 | 4\*10^5 | 4,8\*10^5 | 5\*10^5 |
| Variance | 8,7\*10^10 | 1,7\*10^11 | 1,8\*10^11 | 9\*10^10 | 8\*10^10 | 1,3\*10^10 | 1,7\*10^10 |
| Std Deviation | 2,9\*10^5 | 4,2\*10^5 | 4,2\*10^5 | 3\*10^5 | 3\*10^5 | 3,7\*10^5 | 4,2\*10^5 |
| Minimum | 1,3\*10^5 | 8\*10^4 | 8,5\*10^4 | 7,8\*10^4 | 7,5\*10^4 | 1,2\*10^5 | 1,6\*10^5 |
| Maximum | 1,7\*10^6 | 7,7\*10^6 | 7\*10^6 | 3,6\*10^6 | 3,9\*10^6 | 5,3\*10^6 | 6,9\*10^6 |
| Range | 1,6\*10^6 | 7,6\*10^6 | 7\*10^6 | 3,5\*10^6 | 3,7\*10^6 | 5,2\*10^6 | 7,7\*10^6 |

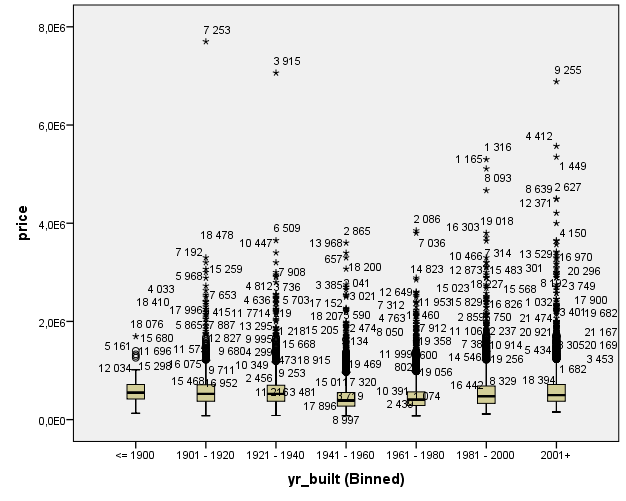
At first glance, it’s doesnt seem to have major price difference between house price by the yr\_built.

**Distribution of prices**



All the distribution are quite similar. So, the distribution of price seems to confirm what we said before, there isn’t have major price difference between house price by the yr\_built. We should use some statistic test to confirm that after.

**Outliers**



With this graphic we can see there is not below outliers. It’s probably because an house can’t be sold below 0.

There are a lot of outliers in absolute value but in percentage it’s low.

Below, the 5 highest outliers for houses price by yr\_built

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  |  | 1 | 2 | 3 | 4 | 5 |
| <1900 | ID | 18076 | 11696 | 15298 | 5161 | 12034 |
| Price | 1,7\*10^6 | 1,4\*10^6 | 1,3\*10^6 | 1 ,3\*10^6 | 1,3\*10^6 |
| 1900-1920 | ID | 7253 | 18478 | 7192 | 15259 | 2900 |
| Price | 7,7\*10^6 | 3,3\*10^6 | 3,2\*10^6 | 3\*10^6 | 3\*10^6 |
| 1920-1940 | ID | 3915 | 6509 | 10447 | 7908 | 11976 |
| Price | 7\*10^6 | 3,7\*10^6 | 3,4\*10^6 | 3,2\*10^6 | 3\*10^6 |
| 1940-1960 | ID | 2865 | 13968 | 18200 | 657 | 14053 |
| Price | 3,6\*10^6 | 3,4\*10^6 | 3,3\*10^6 | 3\*10^6 | 2,8\*10^6 |
| 1960-1980 | ID | 2086 | 7039 | 14557 | 5450 | 13711 |
| Price | 3,9\*10^6 | 3,8\*10^6 | 2,9\*10^6 | 2,9\*10^6 | 2,6\*10^6 |
| 1980-2000 | ID | 1316 | 1165 | 8093 | 19018 | 16303 |
| Price | 6,3\*10^6 | 5,1\*10^6 | 5,7\*10^6 | 3,8\*10^6 | 3,7\*10^6 |
| >2000 | ID | 9255 | 4412 | 1449 | 2627 | 8639 |
| Price | 6.9\*10^6 | 5.6\*10^6 | 5.4\*10^6 | 4.5\*10^6 | 4.5\*10^6 |