**MUBARAK GANIYU**

**ECON 1500-04**

**REAL ESTATE PROJECT**

**02-29-2020**

**Question 1.**

|  |  |
| --- | --- |
| ***Price*** | |
|  |  |
| Mean | 357026.4667 |
| Standard Error | 15682.72727 |
| Median | 323417 |
| Mode | #N/A |
| Standard Deviation | 160700.1342 |
| Sample Variance | 25824533134 |
| Kurtosis | 2.369345513 |
| Skewness | 1.531454313 |
| Range | 751518 |
| Minimum | 167962 |
| Maximum | 919480 |
| Sum | 37487779 |
| Count | 105 |

The range of the prices is about $751518 which suggests that people from different socio-economic background use this real estate agency.

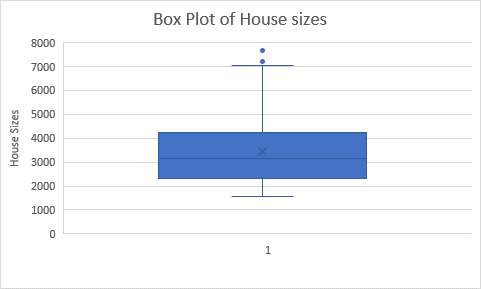
|  |  |
| --- | --- |
| ***Size*** | |
|  |  |
| Mean | 3440.28571 |
| Standard Error | 137.710509 |
| Median | 3150 |
| Mode | 3440 |
| Standard Deviation | 1411.1128 |
| Sample Variance | 1991239.34 |
| Kurtosis | 0.56298267 |
| Skewness | 1.03035297 |
| Range | 6120 |
| Minimum | 1550 |
| Maximum | 7670 |
| Sum | 361230 |
| Count | 105 |

The mean size of the houses is about 3440 sq. feet which is also the most purchased house size i.e. the mode.

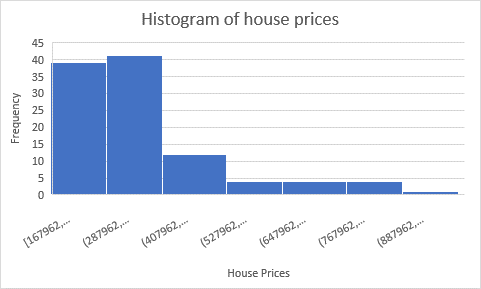
|  |  |
| --- | --- |
| ***FICO*** | |
|  |  |
| Mean | 708.1238 |
| Standard Error | 7.321098 |
| Median | 726 |
| Mode | 813 |
| Standard Deviation | 75.01893 |
| Sample Variance | 5627.84 |
| Kurtosis | -1.3549 |
| Skewness | -0.17215 |
| Range | 241 |
| Minimum | 583 |
| Maximum | 824 |
| Sum | 74353 |
| Count | 105 |

More than 50% of people who purchase houses have a FICO score of above 726 which is greater than the national average FICO score of 706.

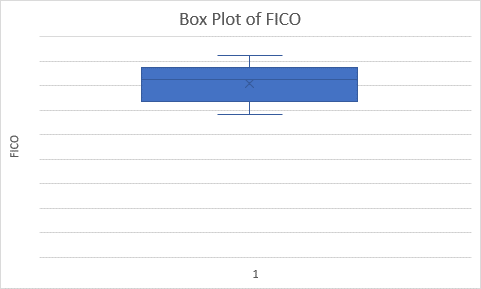
**Question 2.**



The House size data is positively skewed which means that the mode < median < mean.



The house prices are positively skewed which suggests that most of the house prices are below the mean price.



The FICO score data has no eccentric figure which suggests that all the house buyers are within the whiskers.

**Question 3.**

The 95% confidence interval for house price is:

to

We are 95% confident that the house price mean lies in between $325912 to $388141.

The 95% confidence interval for house size is:

to

We are 95% confident that the house size mean lies in between 3167 sq. feet to

3713 sq. feet.

The 95% confidence interval for the FICO score:

to

We are 95% confident that the house buyers’ FICO mean score lies in between 694 sq. feet to 723 sq. feet.

**Question 4.**

The two variables that are likely to have a relationship are price and size. It can be noticed that as size increases, price increases. Moreover, *Inman* website released an article on August 7, 2017 that talks about how sizes play a massive role on housing price. It justifies this claim by linking an article published by *the balance* which proves that there is a relationship between house price and house size. The independent variable will be the house size and the dependent variable will be the house price.

**Question 5.**

|  |
| --- |
|  |
| **CORRELATION COEFFICIENT** |
| 0.95155 |

The correlation coefficient was calculated as 0.95155 which suggests that there is a very strong positive relationship between both variables.

**Question 6.**

**Question 7.**

|  |  |  |
| --- | --- | --- |
| **Size (sq. feet)** | **Price ($)** | **Predicted Price ($)** |
| 1820 | 206424 | 181446.19 |
| 3010 | 346150 | 310399.08 |
| 3210 | 372360 | 332071.84 |
| 3330 | 310622 | 345075.49 |
| 4510 | 496100 | 472944.75 |
| 3440 | 294086 | 356995.51 |
| 2630 | 228810 | 269220.85 |
| 4470 | 384420 | 468610.19 |
| 4040 | 416120 | 422013.77 |
| 4380 | 487494 | 458857.45 |
| 5280 | 448800 | 556384.85 |
| 4420 | 388960 | 463192.01 |
| 2970 | 335610 | 306064.53 |
| 2300 | 276000 | 233460.80 |
| 2970 | 346421 | 306064.53 |
| 3660 | 453913 | 380835.54 |
| 3290 | 376146 | 340740.94 |
| 5900 | 694430 | 623570.39 |
| 2050 | 251269 | 206369.86 |
| 4920 | 547596 | 517373.89 |
| 1950 | 214910 | 195533.48 |
| 1950 | 188799 | 195533.48 |
| 4680 | 459950 | 491366.59 |
| 2540 | 264160 | 259468.11 |
| 3180 | 393557 | 328820.92 |
| 4660 | 478675 | 489199.31 |
| 4220 | 384020 | 441519.25 |
| 3600 | 313200 | 374333.71 |
| 2990 | 274482 | 308231.81 |
| 1920 | 167962 | 192282.57 |
| 1970 | 175823 | 197700.75 |
| 2520 | 226498 | 257300.83 |
| 3150 | 316827 | 325570.01 |
| 1550 | 189984 | 152187.97 |
| 3090 | 366350 | 319068.18 |
| 4080 | 416160 | 426348.32 |
| 3500 | 308000 | 363497.33 |
| 2620 | 294357 | 268137.21 |
| 2790 | 337144 | 286559.05 |
| 2910 | 299730 | 299562.70 |
| 4370 | 445740 | 457773.82 |
| 4200 | 410592 | 439351.98 |
| 5570 | 667732 | 587810.35 |
| 5050 | 523584 | 531461.18 |
| 3360 | 336000 | 348326.40 |
| 2270 | 202598 | 230209.89 |
| 2830 | 326695 | 290893.60 |
| 2770 | 321320 | 284391.78 |
| 2870 | 246820 | 295228.15 |
| 5910 | 546084 | 624654.03 |
| 6800 | 793084 | 721097.79 |
| 1600 | 174528 | 157606.16 |
| 3970 | 392554 | 414428.31 |
| 3060 | 263160 | 315817.27 |
| 1900 | 237120 | 190115.29 |
| 2150 | 225750 | 217206.23 |
| 7190 | 848420 | 763359.67 |
| 3110 | 371956 | 321235.46 |
| 3290 | 404538 | 340740.94 |
| 2810 | 250090 | 288726.33 |
| 3830 | 369978 | 399257.38 |
| 1630 | 209292 | 160857.07 |
| 1850 | 190032 | 184697.10 |
| 2520 | 216720 | 257300.83 |
| 3220 | 323417 | 333155.47 |
| 3070 | 316210 | 316900.91 |
| 2090 | 226054 | 210704.41 |
| 2090 | 183920 | 210704.41 |
| 2300 | 248400 | 233460.80 |
| 5760 | 466560 | 608399.47 |
| 6110 | 667212 | 646326.79 |
| 4370 | 362710 | 457773.82 |
| 3160 | 265440 | 326653.65 |
| 6600 | 706596 | 699425.04 |
| 3300 | 293700 | 341824.58 |
| 2330 | 199448 | 236711.71 |
| 4230 | 369533 | 442602.89 |
| 2030 | 230121 | 204202.58 |
| 1690 | 169000 | 167358.90 |
| 2040 | 190291 | 205286.22 |
| 4660 | 393584 | 489199.31 |
| 2860 | 363792 | 294144.52 |
| 3840 | 360960 | 400341.02 |
| 3180 | 310877 | 328820.92 |
| 7670 | 919480 | 815374.28 |
| 3400 | 392904 | 352660.95 |
| 1840 | 200928 | 183613.46 |
| 4890 | 537900 | 514122.98 |
| 2390 | 258120 | 243213.54 |
| 6160 | 558342 | 651744.98 |
| 3440 | 302720 | 356995.51 |
| 2220 | 240115 | 224791.70 |
| 6530 | 793656 | 691839.57 |
| 1930 | 218862 | 193366.20 |
| 3510 | 383081 | 364580.97 |
| 3380 | 351520 | 350493.68 |
| 7030 | 841491 | 746021.46 |
| 2850 | 336300 | 293060.88 |
| 3750 | 312863 | 390588.28 |
| 3060 | 275033 | 315817.27 |
| 2110 | 229990 | 212871.68 |
| 2130 | 195257 | 215038.96 |
| 1650 | 194238 | 163024.35 |
| 2740 | 348528 | 281140.86 |
| 2240 | 241920 | 226958.97 |

**Question 8.**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | ***Coefficients*** | ***Standard Error*** | ***t Stat*** | ***P-value*** |
| Intercept | -15775.884 | 12821.96 | -1.23038 | 0.221358 |
| Size | 108.363776 | 3.450577 | 31.40454 | 1.46E-54 |

The size coefficient estimate is 108.36 which suggests that for every increase in house size by 1 sq. foot, there is an increase in house price by $108.36. The coefficient estimate signifies that there is an upward slope. The t-stat value for the size coefficient is 31.405 which is greater than the t-critical value of 1.660, which was obtained by setting alpha at 0.05 for a one-tailed t-test distribution. The p-value is 1.46e-54 which is less than 0.05. Thus, the null hypothesis is rejected, the alternate hypothesis accepted and conclude the slope of the line is positive or regression coefficient is greater than zero.

**Question 9.**

The simple linear regression equation can be expressed as:

**Question 10.**

|  |  |
| --- | --- |
| ***Regression Statistics*** | |
| Multiple R | 0.95154564 |
| R Square | 0.9054391 |
| Adjusted R Square | 0.90452103 |
| Standard Error | 49655.8132 |
| Observations | 105 |

The coefficient of determination, r-squared, is 0.905. This suggests that a large proportion of independent variable explains the dependent variable. It also implies that most of the actual house prices are close to the predicted house prices.

**Question 11.**

The residual for the first observation is:

**Question 12.**

The multiple linear regression equation can be expressed as:

**Question 13.**

|  |  |  |  |
| --- | --- | --- | --- |
| **Size (sq. feet)** | **FICO** | **Price ($)** | **Predicted Price ($)** |
| 1820 | 824 | 206424 | 180039.8 |
| 3010 | 820 | 346150 | 309039.7 |
| 3210 | 819 | 372360 | 330724.4 |
| 3330 | 817 | 310622 | 343752.2 |
| 4510 | 816 | 496100 | 471632 |
| 3440 | 813 | 294086 | 355720.7 |
| 2630 | 813 | 228810 | 267947.1 |
| 4470 | 812 | 384420 | 467346.1 |
| 4040 | 810 | 416120 | 420774.6 |
| 4380 | 808 | 487494 | 457642.1 |
| 5280 | 806 | 448800 | 555192.6 |
| 4420 | 805 | 388960 | 462013.1 |
| 2970 | 801 | 335610 | 304936.2 |
| 2300 | 798 | 276000 | 232369.8 |
| 2970 | 795 | 346421 | 305009.1 |
| 3660 | 792 | 453913 | 379815.7 |
| 3290 | 792 | 376146 | 339721.6 |
| 5900 | 788 | 694430 | 622596.1 |
| 2050 | 786 | 251269 | 205425.1 |
| 4920 | 785 | 547596 | 516437.4 |
| 1950 | 784 | 214910 | 194613.2 |
| 1950 | 782 | 188799 | 194637.5 |
| 4680 | 781 | 459950 | 490479.1 |
| 2540 | 780 | 264160 | 258595.6 |
| 3180 | 776 | 393557 | 327996.2 |
| 4660 | 773 | 478675 | 488409.1 |
| 4220 | 772 | 384020 | 440741.7 |
| 3600 | 772 | 313200 | 373557 |
| 2990 | 769 | 274482 | 307492.4 |
| 1920 | 769 | 167962 | 191544.6 |
| 1970 | 766 | 175823 | 196999.2 |
| 2520 | 763 | 226498 | 256635 |
| 3150 | 759 | 316827 | 324952 |
| 1550 | 758 | 189984 | 151584.2 |
| 3090 | 754 | 366350 | 318511 |
| 4080 | 753 | 416160 | 425802 |
| 3500 | 752 | 308000 | 362963.9 |
| 2620 | 751 | 294357 | 267617.1 |
| 2790 | 749 | 337144 | 286063 |
| 2910 | 748 | 299730 | 299078.7 |
| 4370 | 746 | 445740 | 457312.2 |
| 4200 | 741 | 410592 | 438951.3 |
| 5570 | 740 | 667732 | 587420 |
| 5050 | 739 | 523584 | 531083.7 |
| 3360 | 737 | 336000 | 347975.5 |
| 2270 | 737 | 202598 | 229860.5 |
| 2830 | 736 | 326695 | 290555.6 |
| 2770 | 736 | 321320 | 284053.8 |
| 2870 | 735 | 246820 | 294902.2 |
| 5910 | 731 | 546084 | 624372.6 |
| 6800 | 729 | 793084 | 720839.5 |
| 1600 | 728 | 174528 | 157367 |
| 3970 | 726 | 392554 | 414210.3 |
| 3060 | 726 | 263160 | 315600.5 |
| 1900 | 723 | 237120 | 189936.5 |
| 2150 | 715 | 225750 | 217124.4 |
| 7190 | 710 | 848420 | 763331.8 |
| 3110 | 710 | 371956 | 321213.1 |
| 3290 | 707 | 404538 | 340754.8 |
| 2810 | 704 | 250090 | 288777.3 |
| 3830 | 703 | 369978 | 399319.1 |
| 1630 | 701 | 209292 | 160946.1 |
| 1850 | 675 | 190032 | 185101.9 |
| 2520 | 674 | 216720 | 257716.9 |
| 3220 | 673 | 323417 | 333582.7 |
| 3070 | 673 | 316210 | 317328.4 |
| 2090 | 670 | 226054 | 211169.6 |
| 2090 | 669 | 183920 | 211181.8 |
| 2300 | 667 | 248400 | 233962.2 |
| 5760 | 665 | 466560 | 608920.5 |
| 6110 | 662 | 667212 | 646883.9 |
| 4370 | 656 | 362710 | 458406.2 |
| 3160 | 653 | 265440 | 327324.1 |
| 6600 | 652 | 706596 | 700103 |
| 3300 | 647 | 293700 | 342567.8 |
| 2330 | 644 | 199448 | 237492.7 |
| 4230 | 642 | 369533 | 443405.6 |
| 2030 | 639 | 230121 | 205044.7 |
| 1690 | 639 | 169000 | 168201.5 |
| 2040 | 631 | 190291 | 206225.6 |
| 4660 | 630 | 393584 | 490147.3 |
| 2860 | 626 | 363792 | 295143.6 |
| 3840 | 626 | 360960 | 401338.7 |
| 3180 | 624 | 310877 | 329843.9 |
| 7670 | 623 | 919480 | 816403.3 |
| 3400 | 618 | 392904 | 353756.5 |
| 1840 | 618 | 200928 | 184711.1 |
| 4890 | 614 | 537900 | 515265.2 |
| 2390 | 614 | 258120 | 244359.1 |
| 6160 | 613 | 558342 | 652897.6 |
| 3440 | 609 | 302720 | 358200.4 |
| 2220 | 609 | 240115 | 225998.3 |
| 6530 | 605 | 793656 | 693088.9 |
| 1930 | 604 | 218862 | 194633.9 |
| 3510 | 601 | 383081 | 365883 |
| 3380 | 599 | 351520 | 351820.2 |
| 7030 | 596 | 841491 | 747379.6 |
| 2850 | 595 | 336300 | 294436.8 |
| 3750 | 595 | 312863 | 391963 |
| 3060 | 593 | 275033 | 317217.2 |
| 2110 | 591 | 229990 | 214297.2 |
| 2130 | 591 | 195257 | 216464.4 |
| 1650 | 590 | 194238 | 164462.6 |
| 2740 | 584 | 348528 | 282650.6 |
| 2240 | 583 | 241920 | 228481.6 |

**Question 14.**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | ***Coefficients*** | ***Standard Error*** | ***t Stat*** | ***P-value*** |
| Intercept | -7163.51 | 47965.46 | -0.14935 | 0.881574 |
| Size (sq. feet) | 108.3624 | 3.466867 | 31.25658 | 4.71E-54 |
| FICO | -12.1557 | 65.21208 | -0.1864 | 0.852499 |

The size coefficient estimate is 108.36 which suggests that for every increase in house size by 1 sq. foot (given that no change occurs in FICO score), there is an increase in price by $108.36. The FICO coefficient estimate is -12.16 which suggests that for every increase in FICO score by 1 unit (given that no change occurs in size), there is a decrease in price by $12.16.

Reviewing the ANOVA table, it can be inferred that the F test statistic is 488.5 while the F-distribution critical value with an alpha level of 0.05 is approximately 3. The p-value was calculated as 5.67e-53 which is less than 0.05. Therefore, the F stat > F critical and p-value < 0.05 so the null hypothesis is rejected and the alternate hypothesis is accepted. This means that there is at least one of the regression coefficients different from zero.

The t-stat value for the size coefficient is 31.25 is greater than the t-critical value of 1.660, which was obtained by setting alpha at 0.1 for a two tailed t-test distribution. The p-value is 4.71e-54 which is less than 0.05. Thus, reject the null hypothesis and conclude the regression coefficients is different from zero. The t-stat value for the FICO coefficient is -0.186 which is less than the t-critical value of -1.660 by setting alpha at 0.1 for a two tailed t-test distribution. The p-value is 0.852 which is greater than 0.05. Thus, fail to reject the null hypothesis and conclude that the regression coefficient is not different from zero.

**Question 15.**

|  |  |
| --- | --- |
| ***Regression Statistics*** | |
| Multiple R | 0.951563 |
| R Square | 0.905471 |
| Adjusted R Square | 0.903618 |
| Standard Error | 49890.13 |
| Observations | 105 |

The coefficient of determination (with size as the only independent variable), r-squared, is 0.905. When FICO was added as a new independent variable, it is still 0.905. There is no difference between the two r-squared so the fit did not improve. Both coefficients suggest that a large proportion of the independent variables explains the dependent variable. It also implies that most of the actual house prices are close to the predicted house prices.

**Question 16.**

**Heteroskedasticity test:** The residuals value is plotted vs the predicted price it is displayed below:

The variation is residuals tends to increase as the predicted price increases. This suggests that there is a wide gap between the largest and the smallest values in the predicted price values. This results from the fact that the size and price variables have a large range. Thus, this creates a potential issue with heteroskedasticity.

**Multicollinearity test:** The correlation between the FICO variable and the size variable is -0.00209. This is acceptable because it is within the range of -0.7 to 0.7. The r-squared value after setting the FICO as the dependent variable and the size as the independent variable is 4.35e-6. Thus, this leads to a variance inflation factor (VIF) of:

Since the VIF is significantly less than 10, there is no issue with multicollinearity.

**Endogeneity test:** The price of a house can neither influence the size of the house nor the FICO score of the house buyer. Thus, there is no endogeneity. This happens because FICO score is an independent variable that can be used to determine if the person is in a good position to buy the house price does not suggest that the person buying it has a certain FICO score. This same ideology applies to house size. The price of the house does not determine the size. It is the other way around because house size can play a massive role in determining house price but price cannot determine size.

**Autocorrelation test:** A residuals vs predicted price graph was built to test for autocorrelation and its correlation coefficient was calculated.

The graph proves that the relationship between the residuals and the predicted price is non-existent. Thus, there is no pattern to the plotted values. Moreover, the correlation coefficient was 9.3e-16 which suggests that there is no correlation. Thus, there is no issue with autocorrelation.

References:

* <https://www.inman.com/2017/08/07/6-factors-that-influence-a-homes-value/>
* <https://www.thebalance.com/can-i-use-the-price-per-square-foot-to-figure-home-values-1798754>