**ANCOV, House Prices, Geri Bellows**

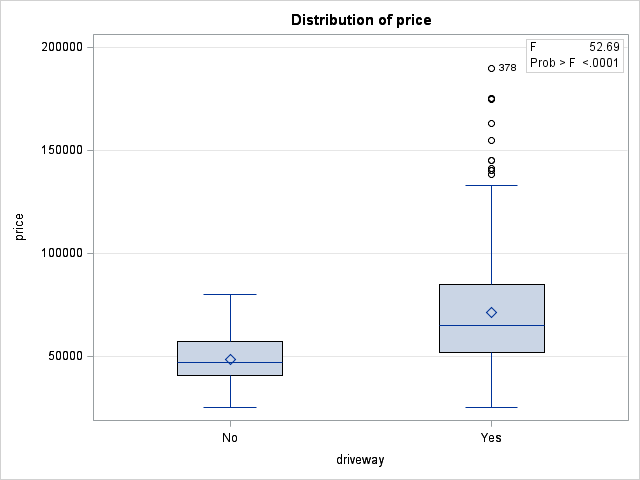
|  |
| --- |
| **Proc** **GLM**; Class driveway; model price=driveway/ss1; means driveway; **run**; **quit**; |

Dependent Variable: price price

| **Source** | **DF** | **Sum of Squares** | **Mean Square** | **F Value** | **Pr > F** |
| --- | --- | --- | --- | --- | --- |
| **Model** | 1 | 34316780536 | 34316780536 | 52.69 | <.0001 |
| **Error** | 544 | 354286005305 | 651261039.16 |  |  |
| **Corrected Total** | 545 | 388602785841 |  |  |  |

| **R-Square** | **Coeff Var** | **Root MSE** | **price Mean** |
| --- | --- | --- | --- |
| 0.088308 | 37.46215 | 25519.82 | 68121.60 |

| **Source** | **DF** | **Type I SS** | **Mean Square** | **F Value** | **Pr > F** |
| --- | --- | --- | --- | --- | --- |
| **driveway** | 1 | 34316780536 | 34316780536 | 52.69 | <.0001 |



OH CRAP ! – The prices appear to more variable and positive skewed in the Yes Driveway group. Let’s ignore that for now.

| **Level of driveway** | **N** | **price** | |
| --- | --- | --- | --- |
| **Mean** | **Std Dev** |
| **No** | **77** | 48555.7792 | 12781.9352 |
| **Yes** | **469** | 71333.8955 | 27027.5772 |

Mean price is $22,778 higher for houses with driveways. That sounds like a big effect to me

If all houses had the same lotsize, would those with driveways still have higher prices? First we test for homogeneity of regression.

|  |
| --- |
| **Proc** **GLM**; Class driveway; model price = lotsize|driveway / ss1; **run**; **quit**; |

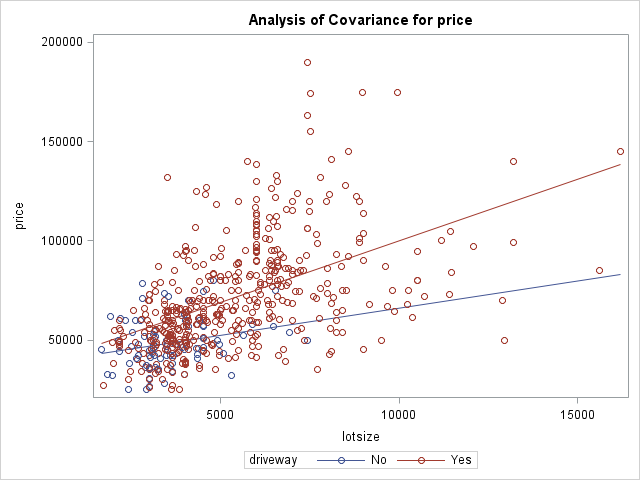
Dependent Variable: price price

| **Source** | **DF** | **Sum of Squares** | **Mean Square** | **F Value** | **Pr > F** |
| --- | --- | --- | --- | --- | --- |
| **Model** | 3 | 121393636341 | 40464545447 | 82.08 | <.0001 |
| **Error** | 542 | 267209149501 | 493005810.89 |  |  |
| **Corrected Total** | 545 | 388602785841 |  |  |  |

| **R-Square** | **Coeff Var** | **Root MSE** | **price Mean** |
| --- | --- | --- | --- |
| 0.312385 | 32.59427 | 22203.73 | 68121.60 |

| **Source** | **DF** | **Type I SS** | **Mean Square** | **F Value** | **Pr > F** |
| --- | --- | --- | --- | --- | --- |
| **lotsize** | 1 | 111558922960 | 111558922960 | 226.28 | <.0001 |
| **driveway** | 1 | 8601857836 | 8601857836 | 17.45 | <.0001 |
| **lotsize\*driveway** | 1 | 1232855544.6 | 1232855544.6 | 2.50 | 0.1144 |

Even with our large sample size, the interaction falls short of significance. The assumption of homogeneity of regression seems reasonable. We drop the interaction from the model.



This plot shows that there is a tendency for price to rise more rapidly with lotsize in houses with driveways than in houses without driveways.

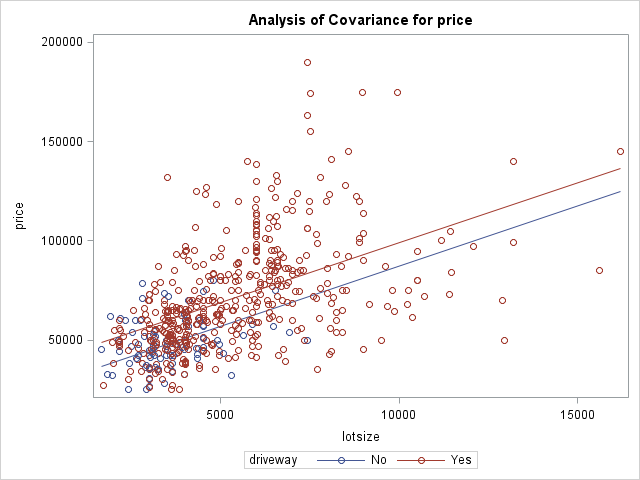
|  |
| --- |
| **Proc** **GLM**; Class driveway; model price = lotsize driveway / ss1; lsmeans driveway; **run**; **quit**; |

Dependent Variable: price price

| **Source** | **DF** | **Sum of Squares** | **Mean Square** | **F Value** | **Pr > F** |
| --- | --- | --- | --- | --- | --- |
| **Model** | 2 | 120160780796 | 60080390398 | 121.53 | <.0001 |
| **Error** | 543 | 268442005045 | 494368333.42 |  |  |
| **Corrected Total** | 545 | 388602785841 |  |  |  |

| **R-Square** | **Coeff Var** | **Root MSE** | **price Mean** |
| --- | --- | --- | --- |
| 0.309212 | 32.63927 | 22234.40 | 68121.60 |

| **Source** | **DF** | **Type I SS** | **Mean Square** | **F Value** | **Pr > F** |
| --- | --- | --- | --- | --- | --- |
| **lotsize** | 1 | 111558922960 | 111558922960 | 225.66 | <.0001 |
| **driveway** | 1 | 8601857836 | 8601857836 | 17.40 | <.0001 |



Since the interaction has been dropped from the model, the within-group slopes are constrained to be identical to each other.

Least Squares Means

| **driveway** | **price LSMEAN** |
| --- | --- |
| **No** | 57889.8644 |
| **Yes** | 69801.4338 |

After controlling for lotsize, houses with driveways are priced $11,911 higher, on average, than those without driveways. When we did not control for lotsize the difference was $22,778.

Remember the problem with heterogeneity of variance and skewness in the one group?

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| |  |  |  | | --- | --- | --- | | **Statisticsa** | | | | price | | | | N | Valid | 469 | | Missing | 0 | | Variance | | 730489930.200 | | Skewness | | 1.134 | | Std. Error of Skewness | | .113 | | Kurtosis | | 1.721 | | Std. Error of Kurtosis | | .225 | | a. driveway = Yes | | | | |  |  |  | | --- | --- | --- | | **Statisticsa** | | | | sr\_Price | | | | N | Valid | 77 | | Missing | 0 | | Variance | | 833.680 | | Skewness | | .148 | | Std. Error of Skewness | | .274 | | Kurtosis | | -.237 | | Std. Error of Kurtosis | | .541 | | a. driveway = No | | | |

Ratio of variances = 730489930/834 = 875887. Oh Crap ! Major heterogeneity of variance. I am going to apply a square root transformation.

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| |  |  |  | | --- | --- | --- | | sr\_Price | | | | N | Valid | 469 | | Missing | 0 | | Variance | | 2343.792 | | Skewness | | .597 | | Std. Error of Skewness | | .113 | | Kurtosis | | .349 | | Std. Error of Kurtosis | | .225 | | a. driveway = Yes | | | | |  |  |  | | --- | --- | --- | | sr\_Price | | | | N | Valid | 77 | | Missing | 0 | | Variance | | 833.680 | | Skewness | | .148 | | Std. Error of Skewness | | .274 | | Kurtosis | | -.237 | | Std. Error of Kurtosis | | .541 | | a. driveway = No | | | |

Hot Damn. Skewness problem resolves. Ratio of variances = 2344/834 = 2.8. Heterogeneity of variances reduced enough not to worry about it anymore.