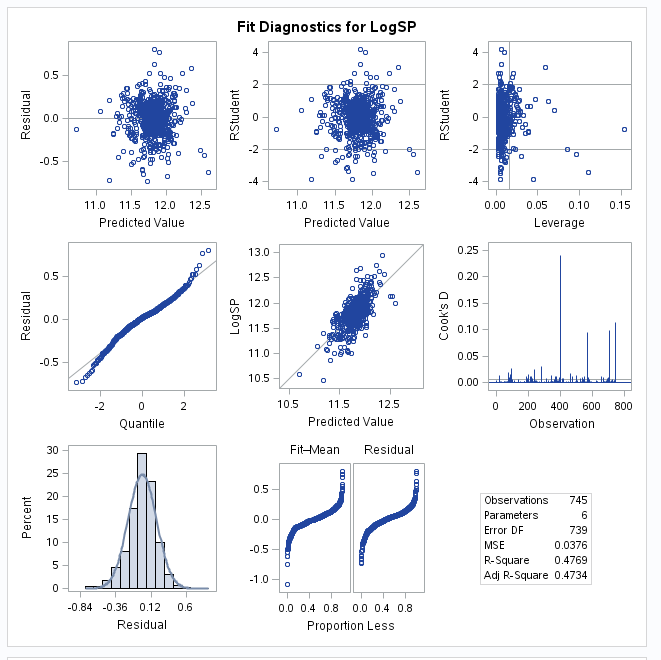
Model:

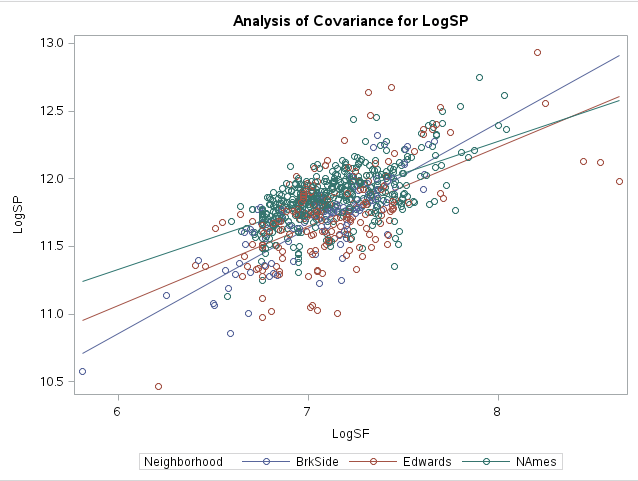
log(Sales Price) = B0 + B1 \* log(Square feet) + B2 \* BrkSide + B3 \* Edwards + B4 \* log(SF) \* BrkSide + B5 \* log(SF) \* Edwards

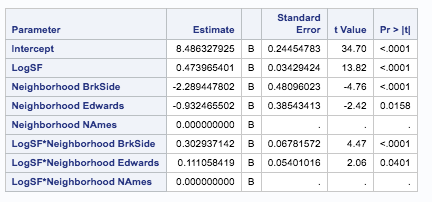
H0: All Bs = 0

Ha: Some Bs != 0

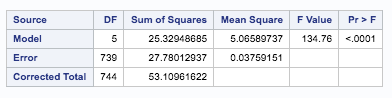


Observations: Looks MOSTLY normal, but there's definitely some curvature in the QQ plot. Maybe the CLT will come into play with 700+ observations? There are a few leverage points, and a couple outliers. I think you'd expect some outliers of that magnitude with the amount of observations we have. Will look at the leverage points later.





All P-values < 0.05 so we want to leave in the interaction terms. Possibly look at the reduced model and do that test only because she had that in her power points lol.



Model:

log(SP) = 8.48 + 0.47 \* log(SF) - 2.29 \* BrkSide - 0.93 \* Edwards + 0.3 \* log(SF) \* BrkSide + 0.11 \* log(SF) \* Edwards

{ log(SP) | BrkSide } = **6.19 + 0.77 \* log(SF)**

{ log(SP) | Edwards } = **7.55 + 0.58 \* log(SF)**

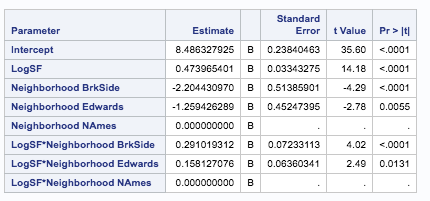
{ log(SP) | NAmes } = 8.48 + 0.47 \* log(SF)

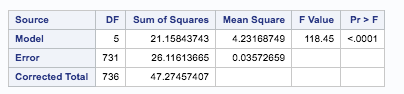
Removed Leverage Points (removed highest 5 and lowest 3):

How do you analyze this?

All variables are still significant once we remove the points.

Intercept almost the same, log(SF) coefficient the same and the other ones are similar but not exactly the same. Is that a good thing?





|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | DF | SS | MS | F | P |
| Model | 8 | 1.66 | 0.2075 | 5.519 | <0.0001 |
| Error | 731 | 26.12 | 0.0376 |  |  |
| Corrected Total | 739 | 27.78 |  |  |  |

Due to increments of 100 sq ft, we may want to use log\_10 and not ln, output for such data, same F and p-values, different coefficients and axes labels:

