Home Prices

Joshua Naranjo
Western Michigan University

Abstract

This study aims to conduct an analysis of factors that determine sale price of homes.

I. Major Comments

1. **Objectives:** Need more focus on the primary goals of the analysis. Is architectural style a significant factor? What are the effect sizes? Is quality of construction a significant factor? What are the effect sizes? What are the confounding variables that need to be adjusted for?
2. **Methods:** One-way ANOVA, two-way ANOVA with interaction, and ANACOVA
3. **Presentation:** Do not underestimate the impact of a well-designed table. It makes your report readable and convincing. Here is Table 4 from one of your reports (the full teport is also posted). I selected this specifically for the design of the efficacy table. You can see the significance and effect sizes across different models. STYLE 7 has highest sale price by itself, but not when SIZE is included in model.

Covariates	Type 3 SS(F-value)	7 >	1 >	3 >	4 >
Style	16.92***	(1, 2, 3, 4, 5)***			
Style	21.55***	(1, 2, 3, 4, 5)***			
Style, Size	11.77***		5**, (6, 7)***	5*, 7***	
Style, Size	10.37***		6**, (5, 7)***	5**, 7***	
Expected Model	11.55***		3**, (2, 7)***		7*
Expected Model	6.16***		2*, 7***		
Full Model	10.82***		3**, (2, 7)***		7**
Full Model	6.95***		2**, 7***		7*
Reduced Model	11.45***		3**, (2, 7)***		7*
Reduced Model	6.91***		2**, 7***		7*

[1] *** = < .001, ** = < .01, * = < .05 represent significant p-values. For Type 3 SS(F-value) this is a partial F-test in SAS. For the last three columns, this is a test for significant differences between lsmeans using tukey's adjustment for multiple comparisons.
[2] The repeated models are the log transformed counterpart.
[3] SSTO =9.81E12, 96.0507 for log transformations.
[4] The expected model consists of the most logically influential variables, using aggregated means and VIF to visually inspect possible confounding. (Style, Size, Quality, Lot Size, and Year)
[5] The reduced model consists of all variables with statistically significant type 3 partial F values. (Style, Size, Quality, Year Garage, Lotsize, Beds, and Highway)

Table 4. Anova of Style on Price (All results)

4. When you change your model, you want to see the effect on significance *and* effect size. (I keep emphasizing this 1-2 punch.) The Type III p-values and LSMEANS do these for you. Whenever a factor or covariate has impact, either on significance or effect size, then this constitues a discovery, and allows you to investigate further.

5. For example, the LSMEANS for STYLE change quite a bit whenever SIZE is in the model (Table 4). SYLE 7 goes from significantly *higher* than (1,4) to signifcantly *lower*. Why? Because STYLE=7 homes average almost 3000 square feet while STYLE 1 or 4 average less than 2000. See Table 1 below.

Table 1: Size of homes (square feet) by STYLE and QUALITY

	QUALITY							
	1		2		3		All	
	N	Mean	N	Mean	N	Mean	N	Mean
STYLE								
1	25	2899.680	77	2083.766	112	1620.652	214	1936.706
4	.	.	4	2583.250	7	1597.857	11	1956.182
7	36	3686.944	94	2774.277	6	2315.000	136	2995.603

6. Table 1 also shows that most STYLE 1 or 4 homes have low quality of construction while STYLE 7 homes have high or middle quality, so need to adjust for QUALITY when reporting means for STYLE. One-way ANOVA versus STYLE alone is misleading, two-way ANOVA versus STYLE and QUALITY is better.