

Multiple Regression, F Tests
COR1-GB.1305 – Statistics and Data Analysis

Multiple Regression

1. We used $n = 294$ from the 2003 Zagat restaurant guide for New York City to fit a regression model, with “Price” as the response variable and “Food,” “Decor,” and “Service” as predictor variables. Here is the output:

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Regression	3	49418.0	16472.7	330.49	0.000
Food	1	19.1	19.1	0.38	0.537
Decor	1	3257.8	3257.8	65.36	0.000
Service	1	5938.5	5938.5	119.14	0.000
Error	290	14454.5	49.8		
Lack-of-Fit	245	12075.7	49.3	0.93	0.640
Pure Error	45	2378.8	52.9		
Total	293	63872.5			

Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)
7.05997	77.37%	77.14%	76.68%

Coefficients

Term	Coef	SE Coef	T-Value	P-Value	VIF
Constant	-20.69	2.31	-8.96	0.000	
Food	-0.103	0.167	-0.62	0.537	2.21
Decor	1.026	0.127	8.08	0.000	2.33
Service	2.555	0.234	10.92	0.000	4.05

Regression Equation

$$\text{Price} = -20.69 - 0.103 \text{ Food} + 1.026 \text{ Decor} + 2.555 \text{ Service}$$

- (a) Interpret the coefficient of “Food” in the context of the estimated multiple regression model. How can this value be negative?
- (b) Does “Food” have utility in explaining “Price” beyond what is explained by “Decor” and “Service”?
- (c) Give a 95% confidence interval for the amount that mean price goes up when we increase food quality rating by 1 point but we hold decor and service ratings constant.

2. In the previous problem, we found that “Food” was not useful for explaining “Price” after adjusting for “Decor” and “Service.” After removing “Food” from the regression model, we get a new regression fit:

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Regression	2	49399	24699.5	496.60	0.000
Decor	1	3802	3802.2	76.45	0.000
Service	1	10586	10586.2	212.84	0.000
Error	291	14474	49.7		
Lack-of-Fit	143	7232	50.6	1.03	0.421
Pure Error	148	7241	48.9		
Total	293	63873			

Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)
7.05247	77.34%	77.18%	76.84%

Coefficients

Term	Coef	SE Coef	T-Value	P-Value	VIF
Constant	-21.39	2.01	-10.63	0.000	
Decor	1.051	0.120	8.74	0.000	2.10
Service	2.455	0.168	14.59	0.000	2.10

Regression Equation

$$\text{Price} = -21.39 + 1.051 \text{ Decor} + 2.455 \text{ Service}$$

Use this regression model to answer the following questions.

- (a) Interpret the coefficient of “Service” in the context of the estimated multiple regression model.
- (b) Does Service have utility in explaining Price beyond what is explained by Decor?
- (c) Give a 95% confidence interval for the amount that mean Price goes up when we increase Service by 1 point but we hold Decor constant.

More Multiple Regression

3. We have a dataset measuring the price (\$), size (ft²), number of bedrooms, and age (years) of 518 houses in Easton, Pennsylvania. We fit a regression model to explain price in terms of the other variables.

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Regression	3	85029785549	28343261850	178.18	0.000
SIZE	1	53484452975	53484452975	336.24	0.000
BEDROOM	1	156773465	156773465	0.99	0.321
AGE	1	279354141	279354141	1.76	0.186
Error	514	81760176401	159066491		
Lack-of-Fit	509	80933266401	159004453	0.96	0.607
Pure Error	5	826910000	165382000		
Total	517	1.66790E+11			

Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)
12612.2	50.98%	50.69%	50.19%

Coefficients

Term	Coef	SE Coef	T-Value	P-Value	VIF
Constant	25875	3555	7.28	0.000	
SIZE	39.20	2.14	18.34	0.000	1.71
BEDROOM	-1145	1153	-0.99	0.321	1.71
AGE	-354	267	-1.33	0.186	1.01

Regression Equation

$$\text{PRICE} = 25875 + 39.20 \text{ SIZE} - 1145 \text{ BEDROOM} - 354 \text{ AGE}$$

- Do the signs of the coefficients make sense to you? Explain any apparent contradictions between what you would expect and what the Minitab output indicates.
- What does the result of the t test on the coefficient of Size indicate?
- What does the result of the t test on the coefficient of Bedroom indicate?
- What does the result of the regression F test indicate?

Regression F Tests

4. Locate the regression F statistic and the corresponding p value in the output from the previous problem.
 - (a) How is the regression F statistic computed?
 - (b) How many numerator and denominator degrees of freedom are there in the regression F statistic?
 - (c) How is the p -value computed?
 - (d) What are the null and alternative hypothesis for the regression F test?
 - (e) Based on the p -value, what is the conclusion of the regression F test (use a significance level of 5%)?