

Regression with Qualitative and Quantitative Variables
STAT-UB.0003 – Regression and Forecasting Models

Multiple Regression with Qualitative Predictors (Review)

1. We asked 46 NYU students how much time they spend on social media, and what their primary computer is (Mac or PC). We are going to use regression to find out if one type of computer associated is with more social media usage. We have the response variable

Social = amount of time (in minutes per week) using social media

We would like to use “OS” as a predictor variable, which is a categorical (qualitative) variable taking values in the set {Mac, PC}.

- (a) How can we encode the OS qualitative variable in terms of one or more quantitative variables?
 - (b) Give a model that relates OS to Social media usage, using an intercept term and a dummy variable for “PC”.
 - (c) What is the interpretation of β_0 and β_1 ?
2. Using the data from problem 1, we fit the regression model in Minitab, and got the following output.

Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)
285.436	5.28%	3.13%	0.00%

Coefficients

Term	Coef	SE Coef	T-Value	P-Value	VIF
Constant	295.2	57.1	5.17	0.000	
OS_PC	-132.3	84.5	-1.57	0.124	1.00

Regression Equation

Social = 295.2 - 132.3 OS_PC

- (a) What is the estimated mean social usage for Mac users?
- (b) What is the estimated mean social usage for PC users?
- (c) What is the interpretation of the p-value for the test on the coefficient of PC?

3. We use the same data, but now we are interested in whether or not texting behavior differs by cell phone type (Blackberry, iPhone, other smart phone, or standard cell phone).

(a) Introduce dummy variables to encode cell phone type.

(b) Using the variables you defined in part (a), devise a regression model which explains text usage in terms of cell phone type.

(c) What is the interpretation of β_0 , the intercept?

(d) What are the interpretations of the other coefficients in your model?

4. We fit a model that explains Text in terms of cell phone type using dummy variables for cell phone type.

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Regression	3	1025437	341812	0.57	0.640
Cell_Blackberry	1	19802	19802	0.03	0.857
Cell_iPhone	1	584505	584505	0.97	0.330
Cell_Smartphone	1	18678	18678	0.03	0.861
Error	42	25299274	602364		
Total	45	26324711			

Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)
776.121	3.90%	0.00%	0.00%

Coefficients

Term	Coef	SE Coef	T-Value	P-Value	VIF
Constant	132	317	0.42	0.680	
Cell_Blackberry	91	501	0.18	0.857	1.52
Cell_iPhone	349	354	0.99	0.330	2.39
Cell_Smartphone	68	388	0.18	0.861	2.22

Regression Equation

$$\text{Text} = 132 + 91 \text{ Cell_Blackberry} + 349 \text{ Cell_iPhone} + 68 \text{ Cell_Smartphone}$$

- (a) What is the estimated mean Text usage for people without smart phones?
- (b) What is the estimated mean Text usage for people with iPhones?
- (c) Is there statistically significant evidence that people with iPhones exhibit different texting behavior (volume) than people without smart phones?
- (d) Is cell phone type useful for predicting Text?

Multiple Regression with Qualitative and Quantitative Predictors

5. Suppose we want to explain Social (minutes per week) in terms of OS (PC or Mac) and Email (minutes per week). Here is the regression output:

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Regression	2	390597	195299	2.47	0.096
OS_PC	1	293693	293693	3.72	0.060
Email	1	190702	190702	2.42	0.127
Error	43	3394150	78934		
Lack-of-Fit	29	2762459	95257	2.11	0.071
Pure Error	14	631692	45121		
Total	45	3784748			

Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)
280.951	10.32%	6.15%	0.64%

Coefficients

Term	Coef	SE Coef	T-Value	P-Value	VIF
Constant	249.0	63.6	3.92	0.000	
OS_PC	-165.7	85.9	-1.93	0.060	1.07
Email	0.729	0.469	1.55	0.127	1.07

Regression Equation

$$\text{Social} = 249.0 - 165.7 \text{ OS_PC} + 0.729 \text{ Email}$$

- Interpret the estimated regression coefficients in the context of the model.
- Interpret the p-value for the coefficient of PC.
- Interpret the p-value for the coefficient of Email.
- Interpret the p-value for the ANOVA F test.
- What assumptions to the various regression hypothesis tests rely on?