Name: Libin Guo

Course: Regression Analysis

Submission Date: January 2nd, 2015

Session: Fall 2010 VEE Student Project

Regression Analysis for Marriage Rate

I. Introduction

The marriage rate in America has declined gradually. There are many variables that could help explain the phenomena. Some of such variables are employment rate of men, female's educational attainment rate for a bachelor's or higher degree, and cohabitation rate of unmarried couples. This paper would conduct a regression analysis of marriage rate by using these three selected variables.

II. Data

The variables are defined as:

Y = Marriage Rate
X_1 = MenEmplymentRate
X_2 = FemaleBachelorDegreeRate
X_3 = CohabitationRateOfUnmarried

The resulting regression model with all explanatory variables is:

$$Y = \alpha + ß_1 X_1 + ß_2 X_2 + ß_3 X_3$$

The data is obtained as follows:

- Marriage Rate: Centers for Disease Control and Prevention,
  - http://www.cdc.gov/nchs/nvss/marriage_divorce_tables.htm
- MenEmploymentRate: OECD Labor Force Statistics, LFS by sex and age – indicators
  - http://stats.oecd.org/index.aspx?queryid=254#
- FemaleBachelorDegreeRate: National Center for Education Statistics
  - http://nces.ed.gov/programs/digest/d13/tables/dt13_104.20.asp
- CohabitationRateofUnmarried: Family Facts Organization
  - http://www.familyfacts.org/charts/110/nearly-12-percent-of-couples-living-together-are-unmarried

A summary of data is displayed in Appendix A.

III. Regression Analysis

3-Variable Regression

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.9754
R Square	0.9514
Adjusted R Square	0.9332
Standard Error	0.1358
Observations	12

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	3	2.8891	0.9630	52.2214	0.0000
Residual	8	0.1475	0.0184		
Total	11	3.0367			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	2.5519	7.5547	0.3378	0.7442	-14.8693	19.9731	-14.8693	19.9731
MenEmplymentRate	0.1143	0.0739	1.5467	0.1605	-0.0561	0.2846	-0.0561	0.2846
FemaleBachelorDegreeRate	-0.1259	0.0996	-1.2646	0.2416	-0.3556	0.1037	-0.3556	0.1037
CohabitationRateOfUnmarried	-0.0350	0.1150	-0.3041	0.7688	-0.3002	0.2302	-0.3002	0.2302

The R^2 with all explanatory variables in the model is 0.9514 and adjusted R^2 is slightly lower at 0.9332, indicating that 93.32% of the variance in the marriage rate can be explained by the 3 variables in the model. The result shows that MenEmploymentRate is positively correlated with marriage rate, while FemaleBachelorDegreeRate and CohabitationOfUnmarried are negatively correlated with marriage rate. Some study has shown that more women in America are choosing to live with men first without marriage, indicating a higher cohabitation rate might be a factor negatively impacting marriage rate. However it's high P-value suggests no predictive power. The variable will be excluded in the next analysis.

2-Variable Regression

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.9751
R Square	0.9509
Adjusted R Square	0.9399
Standard Error	0.1288
Observations	12

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	2	2.8874	1.4437	87.0651	0.0000
Residual	9	0.1492	0.0166		
Total	11	3.0367			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	2.9365	7.0626	0.4158	0.6873	-13.0403	18.9132	-13.0403	18.9132
MenEmplymentRate	0.1165	0.0697	1.6703	0.1292	-0.0413	0.2742	-0.0413	0.2742
FemaleBachelorDegreeRate	-0.1526	0.0447	-3.4117	0.0077	-0.2538	-0.0514	-0.2538	-0.0514

In the 2-variable model, the R^2 has slightly decreased, however the adjusted R^2 has slightly increased. In addition, there is improvement in F and standard error indicator. F has increased significantly from 52.2214 to 87.0651 and standard error has decreased from 0.1358 to 0.1288. Thus the 2-variable model is better than 3-variable model.

The P-value for MenEmploymentRate appears to be higher than 5% at 12.92%, indicating the variable is not so significant in t-Stat. The variable MenEmploymentRate fall out of significance at the 95% level and hence will be removed in the next analysis.

1-Variable Regression

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.9673
R Square	0.9356
Adjusted R Square	0.9292
Standard Error	0.1398
Observations	12

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	1	2.8412	2.8412	145.3277	0.0000
Residual	10	0.1955	0.0196		
Total	11	3.0367			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	14.6971	0.5998	24.5052	0.00000000	13.3608	16.0335	13.3608	16.0335
FemaleBachelorDegreeRate	-0.2217	0.0184	-12.0552	0.00000028	-0.2627	-0.1808	-0.2627	-0.1808

It is noticed that the adjusted R^2 has decreased from 0.9399 to 0.9292, and standard error has increased slightly from 0.1288 to 0.1398. This means that the model is slightly worse. However the F value, which indicates the

ratio of explained variance to unexplained variance, has been increased significantly from 87.0651 to 145.3277 and the variable is statically significant at the 95% level, with a P-value close to zero.

IV. Conclusion

In the 1-variable model, adjusted R^2 is 0.9292 which means about 92.92% of the variation in the response variable can be explained by FemaleBachelrDegreeRate. In other words, the predictive ability for this variable is well sufficient in that there exists only about 7% of the variation in the response cannot be explained. Although the 2-variable model seems to be a slightly better one based on the slightly higher adjusted R^2 value, the significant level of the variables and number of variables should be taken into consideration when selecting a final model. The 1-variable model is preferred over all models, since the model is a good fit with a high adjusted $R^{2,}$ less variables and the significant level of the variable being above 95%.

The regression model is $Y = 14.6971 - 0.2217X_2$

Y = Marriage Rate
X_2 = FemaleBachelorDegreeRate

Appendix A - Summary of data

	Y	X_1	X_2	X_3
Year	Marriage Rate (Rate per 1,000 total population)	MenEmploymentRate (Men 15to64 Labor force participation rate)	FemaleBachelorDegreeRate (Female Bachelor's or higher degree attainment rate)	CohabitationRateOfUnmarried (Cohabiting couples as a % of all coresidential couples)
2000	8.2	83.9	30.1	6.5
2001	8.2	83.4	30.1	6.8
2002	8	83	30.1	6.9
2003	7.7	82.2	30.9	7.5
2004	7.8	81.9	30.9	7.5
2005	7.6	81.8	32.2	7.8
2006	7.5	81.9	31.6	7.9
2007	7.3	81.7	33	9.9
2008	7.1	81.4	34.9	10.5
2009	6.8	80.4	34.8	10.1
2010	6.8	79.6	35.7	11.4
2011	6.8	78.9	36.1	11.6

Y- Centers for Disease Control and Prevention, http://www.cdc.gov/nchs/nvss/marriage_divorce_tables.htm

X_1 - OECD Labor Force Statistics LFS by sex and age - indicators, http://stats.oecd.org/index.aspx?queryid=254#

X_2 - National Center for Education Statistics, http://nces.ed.gov/programs/digest/d13/tables/dt13_104.20.asp

X_3 - Family Facts Organization, http://www.familyfacts.org/charts/110/nearly-12-percent-of-couples-living-together-are-unmarried