

Divorce rates in the US

Introduction

The divorce rates in the US have been rising consistently for the past 50 years. Divorce rates are the lowest in Massachusetts at 1.8 per 1000 of population and the highest in Nevada at 6.6 per 1000 of population. In the study below I will try to find a predictor for the divorce rates.

Data

The data for the analysis was collected from the following web sites and is covering all of the US states, excluding Hawaii:

www.census.gov

www.americanreligionsurvey-aris.org

www.well-beingindex.com

The response variable used in the models is number of divorces per 1000 of population.

Y: Number of Divorces in a State (per 1000 of population)

The explanatory variables used in the models are:

X₁: Median state family income (in \$1000)

X₂: Statutory unemployment rate

X₃: Percentage of people that state religious observance

X₄: State Education Ranking Index (SERI)

X₅: Well-being Index as per well-beingindex.com site

X₆: Number of marriages in a state (per 1000 of population)

Multi-collinearity

To begin, I would like to explore correlations between the explanatory variables. Below is the summary of the correlations between the explanatory variables.

	<i>Divorce (per 1000)</i>	<i>Median Family Income (in \$1000)</i>	<i>Unempl</i>	<i>% Religious</i>	<i>SERI</i>	<i>Well- being Index</i>	<i>Mariage (per 1000)</i>
Divorces (per 1000)	1.00						
Median Family Income (in \$1000)	0.02	1.00					
Unemployment	0.21	0.18	1.00				
% Religious	-0.14	-0.35	0.07	1.00			
SERI	-0.47	0.11	-0.16	-0.31	1.00		
Well-being Index	-0.35	-0.08	-0.50	-0.32	0.44	1.00	
Mariages (per 1000)	-0.17	-0.11	-0.22	-0.09	0.04	0.23	1.00

Few observations to note from the table above:

- Median State family income has little correlation with the divorce rate.
- Unemployment is positively correlated with the divorce rate.
- SERI is moderately negatively correlated with the divorce rate, i.e. the higher SERI, the lower divorce rate.
- Similarly, well-being index is moderately correlated with the divorce rate, i.e. the higher the well-being index, the lower divorce rate.
- The well-being index is positively correlated with SERI, implying that the better people educated, the happier they are.
- There are no obvious strong correlations between Divorce rate and explanatory variables. This may suggest poor choice of explanatory variables.

Model 1

The model uses six explanatory variables resulting in the regression equation below:

$$Y = 18.60 - 0.01 * X_1 + 1.69 * X_2 - 0.06 * X_3 - 0.58 * X_4 - 0.12 * X_5 - 0.08 * X_6$$

<i>Regression Statistics</i>	
Multiple R	0.63
R Square	0.40
Adjusted R Square	0.31
Standard Error	0.74
Observations	50

	<i>Coefficients</i>	<i>Standard Error</i>	<i>t Stat</i>	<i>P-value</i>
Intercept	18.60	6.40	2.91	0.006
Median Family Income (in	-0.01	0.02	-0.83	0.409

\$1000)				
Unemployment	1.69	6.32	0.27	0.790
Percent of Religious population	-0.06	0.02	-2.98	0.005
SERI	-0.58	0.17	-3.45	0.001
Well-being Index	-0.12	0.08	-1.41	0.166
Marriages (per 1000)	-0.08	0.07	-1.14	0.262

This model appears to be not a great fit, as illustrated by the R square value of 0.40. The two highest P-values are 0.79 for unemployment rate and 0.41 for median family income. These values indicate that variables above have the least predictive power when it comes to predicting divorce rates. For my next model I will remove X_1 and X_2 variables.

Model 2

The model uses four explanatory variables resulting in the regression equation below:

$$Y = 17.51 - 0.05*X_3 - 0.59*X_4 - 0.12X_5 - 0.08*X_6$$

<i>Regression Statistics</i>	
Multiple R	0.62
R Square	0.39
Adjusted R Square	0.33
Standard Error	0.73
Observations	50

	<i>Coefficients</i>	<i>Standard Error</i>	<i>t Stat</i>	<i>P-value</i>
Intercept	17.51	5.08	3.45	0.0013
Percent of Religious population	-0.05	0.02	-2.95	0.0051
SERI	-0.59	0.16	-3.58	0.0009
Well-being Index	-0.12	0.07	-1.60	0.1180
Marriages (per 1000)	-0.08	0.07	-1.11	0.2719

Compared to Model 1, this model is a worse fit since R-square value is slightly lower. However, since we reduced the degrees of freedom and did not significantly lose on the R-square value, the removal of variables is warranted. The next variable we will try to remove X_6 variables, i.e. Marriages by State per 1000 of population. This variable has P-value of .11 and is not a significant predictor.

Model 3

The model uses three explanatory variables resulting in the regression equation below:

$$Y = 18.15 - 0.05*X_3 - 0.57*X_4 - 0.14*X_5$$

<i>Regression Statistics</i>	
Multiple R	0.61
R Square	0.37
Adjusted R Square	0.33
Standard Error	0.73
Observations	50

	<i>Coefficients</i>	<i>Standard Error</i>	<i>t Stat</i>	<i>P-value</i>
Intercept	18.15	5.06	3.59	0.0008
Percent of Religious population	-0.05	0.02	-2.91	0.0056
SERI	-0.57	0.16	-3.49	0.0011
Well-being Index	-0.14	0.07	-1.89	0.0649

As in Model 2, R-square has decreased but the tradeoff between reducing the degrees of freedom and marginally decreasing the R-square value of warranted.

Conclusion:

The third model shows that the most important factor in predicting the divorce rates in a State is SERI, the State Education Ranking Index. This implies that more educated people have a lower rate of divorce. It may or may not be the case in practice as it is difficult to believe that more education would make people better husbands and wives. The other two most important predictors are percent of religious population and a well-being index. The greater the State's religious population, the lower the rate of divorce in the State. That being said, with some exceptions, Christian belt states have much lower divorce rates than the rest of the states. Lastly, as expected, the Well-being Index could predict rate of divorces in a State – the happier the people are in a State, the less likely they are to divorce.

The three explanatory variables outlined in the third model are quite subjective in nature. The R-square of 0.33 implies poor goodness of fit of the model to the actual world.