VEE Fall 2014 Regression Analysis Student Project

Birth rate in Taiwan

Yen-Chen Lin

Introduction

In Taiwan, the birth rate declined gradually . So I would like to know the relationship between the birth rate in Taiwan and some factors such as marriage rate, CPI, Economic growth rate and so on.

Data and Consideration

The data was obtained from National Statistics, R.O.C.(Taiwan). I was able to obtain the statistic for all factors between year 1994 and 2013.(http://www.stat.go.jp/index.htm)

The initial model will incorporate eight explanatory variables (seven independent variables and one constant term). I used the Excel Toolpak to do regression analysis. Subsequent models will be improved upon by removing the variable with the highest p-value, indicating that variable has the lowest predictive power. I would observe the Adjusted R Square values as opposed to just R Square value as they are not sensitive to the number of variables. This process will be repeated until I can obtain the most predictive model for determining the key factors that explain the birth rate in Taiwan.

Regression equation:

$Y = \alpha + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \beta_5 X_5 + \beta_6 X_6 + \beta_7 X_7$

Where α = Intercept,

βi = Slope for variable I,

Y = Birth Rate % per thousand,

X_1 = Marriage Rate % per thousand,

X_2 = Consumer Price Index (General),

X_3 = Consumer Price Index (Education),

X_4 = Consumer Price Index (Food),

X_5 = Average Income,

X_6 = Economic Growth Rate %,

X_7 = Unemployment rate %,

Initial Model:

$Y = -12.8285 + 0.4048 X_1 + 1.1993 X_2 - 0.2768 X_3 - 0.6640 X_4 - 0.0002 X_5 - 0.1158 X_6 - 0.6136 X_7$

SUMMARY OUT

Regression Statistics	
Multiple R	0.9736

	R Square	0.9480
	Adjusted R Square	0.9176
	Standard Error	0.7867
	Observations	20.0000

ANOVA

	DF	SS	MS	F	Significance F
Regression	7	135.3242	19.3320	31.2356	0.0000
Residual	12	7.4269	0.6189		
Total	19	142.7512			

	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%
Intercept	-12.8285	17.7766	-0.7216	0.4843	-51.5603	25.9034
Marriage Rate	0.4048	0.3353	1.2071	0.2506	-0.3258	1.1354
CPI (General)	1.1993	0.5125	2.3403	0.0374	0.0828	2.3159
CPI (Education)	-0.2768	0.0885	-3.1267	0.0087	-0.4697	-0.0839
CPI (Food)	-0.6640	0.2538	-2.6159	0.0226	-1.2171	-0.1110
Average Income	-0.0002	0.0003	-0.5579	0.5872	-0.0010	0.0006
Economic Growth Rate	-0.1158	0.0852	-1.3588	0.1992	-0.3015	0.0699
Unemployment Rate	-0.6136	0.3303	-1.8575	0.0879	-1.3333	0.1061

This model has a high Adjusted R Square value of 91.76%, indicating that 91.76% of the variance in the birth rate is explained by the 7 variables in the model. Based on the results, Average Income has the highest P-value. The high P-value indicate that this may not be a good explanatory variable. So I remove Average Income and regress again.

Model 2 with 6 variables:

$Y = -7.8690+0.3732X_1+1.1111X_2-0.2685X_3-0.6659X_4-0.1449X_5-0.5770X_6$

SUMMARY OUT

Regression Statistics	
Multiple R	0.9729
R Square	0.9466
Adjusted R Square	0.9220
Standard Error	0.7656
Observations	20.0000

ANOVA

	DF	SS	MS	F	Significance

				F
Regression	6	135.1316	22.5219	38.4253	0.0000
Residual	13	7.6196	0.5861		
Total	19	142.7512			

	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%
Intercept	-7.8690	14.9810	-0.5253	0.6082	-40.2335	24.4955
Marriage Rate	0.3732	0.3216	1.1603	0.2668	-0.3217	1.0680
CPI (General)	1.1111	0.4744	2.3422	0.0357	0.0863	2.1359
CPI (Education)	-0.2685	0.0849	-3.1616	0.0075	-0.4520	-0.0850
CPI (Food)	-0.6659	0.2470	-2.6959	0.0183	-1.1995	-0.1323
Economic Growth Rate	-0.1449	0.0657	-2.2050	0.0461	-0.2868	-0.0029
Unemployment Rate	-0.5770	0.3151	-1.8314	0.0901	-1.2576	0.1037

The adjusted R Square has improved to 92.20%. Marriage Rate has the highest P-value. I remove Marriage Rate and regress again.

Model 3 with 5 variables:
$$Y = -12.8852 + 1.3841X_1 - 0.3099X_2 - 0.8207X_3 - 0.1562X_4 - 0.6269X_5$$

SUMMARY OUT

Regression Statistics	
Multiple R	0.9701
R Square	0.9411
Adjusted R Square	0.9201
Standard Error	0.7750
Observations	20.0000

ANOVA

	DF	SS	MS	F	Significance F
Regression	5	134.3425	26.8685	44.7345	0.0000
Residual	14	8.4087	0.6006		
Total	19	142.7512			

	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%
Intercept	-12.8852	14.5200	-0.8874	0.3898	-44.0275	18.2570
CPI (General)	1.3841	0.4170	3.3194	0.0051	0.4898	2.2785
CPI (Education)	-0.3099	0.0780	-3.9725	0.0014	-0.4772	-0.1426

	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%
CPI (Food)	-0.8207	0.2104	-3.9010	0.0016	-1.2720	-0.3695
Economic Growth Rate	-0.1562	0.0658	-2.3761	0.0323	-0.2973	-0.0152
Unemployment Rate	-0.6269	0.3159	-1.9844	0.0672	-1.3046	0.0507

The adjusted R Square was 92.01%. Unemployment Rate has the P-value>0.05. So I remove Unemployment Rate and regress again.

Model 4 with 4 variables:
$$Y = -16.5129+1.6069X_1-0.4027X_2-0.9419X_3-0.1384X_4$$

SUMMARY OUT

Regression Statistics	
Multiple R	0.9615
R Square	0.9245
Adjusted R Square	0.9044
Standard Error	0.8475
Observations	20.0000

ANOVA

	DF	SS	MS	F	Significance F
Regression	4	131.9774	32.9944	45.9372	0.0000
Residual	15	10.7737	0.7182		
Total	19	142.7512			

	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%
Intercept	-16.5129	15.7519	-1.0483	0.3111	-50.0874	17.0615
CPI (General)	1.6069	0.4391	3.6591	0.0023	0.6709	2.5429
CPI (Education)	-0.4027	0.0683	-5.8957	0.0000	-0.5483	-0.2571
CPI (Food)	-0.9419	0.2202	-4.2779	0.0007	-1.4112	-0.4726
Economic Growth Rate	-0.1384	0.0712	-1.9428	0.0711	-0.2902	0.0134

The adjusted R Square was 90.44%. Economic Growth Rate has the P-value>0.05. So I remove Economic Growth Rate and regress again.

Model 5 with 3 variables:
$$Y = -16.4387+1.5354X_1-0.3823X_2-0.8966X_3$$

SUMMARY OUT

Regression Statistics	
Multiple R	0.9516
R Square	0.9055
Adjusted R Square	0.8878
Standard Error	0.9180
Observations	20.0000

ANOVA

	DF	SS	MS	F	Significance F
Regression	3	129.2664	43.0888	51.1258	0.0000
Residual	16	13.4848	0.8428		
Total	19	142.7512			

	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%
Intercept	-16.4387	17.0631	-0.9634	0.3497	-52.6108	19.7334
CPI (General)	1.5354	0.4740	3.2390	0.0051	0.5305	2.5403
CPI (Education)	-0.3823	0.0731	-5.2293	0.0001	-0.5373	-0.2273
CPI (Food)	-0.8966	0.2372	-3.7805	0.0016	-1.3994	-0.3938

Although the Adjusted R Square reduced from 90.44% to 88.78%, F value has increased to 51.1 and the variables are statistically significant at 95% level.

Conclusion

I would choose the model 5 with three variables because I can get the good fit model by using less variables. And the model is not purely on Adjusted R square but also at the significant level of the variables.

So I conclude that the CPI (general), CPI (education) and CPI (Food) is a good predictor of birth rate. CPI (general), CPI (education) and CPI (Food) has a negative impact on the birth rate, so I suspect that higher cost would lead to lower inclination for family to have more babies.

Data

Year	Birth Rate % per thousand	Marriage Rate % per thousand	Consumer Price Index (General)	Consumer Price Index (Education)	Consumer Price Index (Food)
1994	15.31	8.1	80.77	73.32	71.18
1995	15.5	7.53	83.73	77.5	74.25
1996	15.18	7.9	86.31	84.84	77.08
1997	15.07	7.68	87.09	88.73	76.54

1998	12.43	6.69	88.56	92.27	79.91
1999	12.89	7.87	88.71	95.29	79.33
2000	13.76	8.19	89.82	99.01	79.64
2001	11.65	7.63	89.82	102.71	78.93
2002	11.02	7.69	89.64	103.52	78.79
2003	10.06	7.6	89.39	102.51	78.73
2004	9.56	5.8	90.83	102.13	82.05
2005	9.06	6.21	92.92	102.32	88
2006	8.96	6.25	93.48	102.53	87.43
2007	8.92	5.89	95.16	102.62	89.93
2008	8.64	6.73	98.51	102.47	97.62
2009	8.29	5.07	97.66	101.33	97.19
2010	7.21	6	98.6	100.13	97.79
2011	8.48	7.13	100	100	100
2012	9.86	6.16	101.93	99.78	104.16
2013	8.53	6.32	102.74	99.75	105.48

Year	Average Income(USD)	Economic Growth Rate %	Unemployment rate %
1994	11092	7.49	1.56
1995	11909	6.5	1.79
1996	12399	6.18	2.6
1997	12701	6.11	2.72
1998	11495	4.21	2.69
1999	12330	6.72	2.92
2000	13341	6.42	2.99
2001	11936	-1.26	4.57
2002	12270	5.57	5.17
2003	12642	4.12	4.99
2004	13735	6.51	4.44
2005	14602	5.42	4.13
2006	14974	5.62	3.91
2007	15401	6.52	3.91
2008	15388	0.7	4.14
2009	14398	-1.57	5.85
2010	16650	10.63	5.21
2011	17982	3.8	4.39
2012	18125	2.06	4.24
2013	18700	2.23	4.18