Reppeto530Week9

February 11, 2024

0.1 Chapter 11

Brian Reppeto 530 Prof. Jim Week 9 HW 11-1

[6]:
```
# import files and libraries

import first
live, firsts, others = first.MakeFrames()
live = live[live.prglngth>30]
import statsmodels.formula.api as smf
import nsfg
import numpy as np
import pandas as pd
```

[7]:
```
# performs an OLS regression analysis to explore the relationship between the␣
 ↪length of
# pregnancy (prglngth) and the variables birthord, race, and nbrnaliv in the␣
 ↪live dataset

model = smf.ols('prglngth ~ birthord==1 + race==2 + nbrnaliv>1', data=live)
results = model.fit()
results.summary()
```

[7]:

Dep. Variable:	prglngth	**R-squared:**	0.011
Model:	OLS	**Adj. R-squared:**	0.011
Method:	Least Squares	**F-statistic:**	34.28
Date:	Sat, 10 Feb 2024	**Prob (F-statistic):**	5.09e-22
Time:	16:10:36	**Log-Likelihood:**	-18247.
No. Observations:	8884	**AIC:**	3.650e+04
Df Residuals:	8880	**BIC:**	3.653e+04
Df Model:	3		
Covariance Type:	nonrobust		

	coef	**std err**	**t**	**P> |t|**	**[0.025**	**0.975]**
Intercept	38.7617	0.039	1006.410	0.000	38.686	38.837
birthord == 1[T.True]	0.1015	0.040	2.528	0.011	0.023	0.180
race == 2[T.True]	0.1390	0.042	3.311	0.001	0.057	0.221
nbrnaliv > 1[T.True]	-1.4944	0.164	-9.086	0.000	-1.817	-1.172

Omnibus:	1587.470	**Durbin-Watson:**	1.619
Prob(Omnibus):	0.000	**Jarque-Bera (JB):**	6160.751
Skew:	-0.852	**Prob(JB):**	0.00
Kurtosis:	6.707	**Cond. No.**	10.9

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.

11-3

[8]:
```
# filter the live df to include only pregnancies with a length greater than 30 weeks

live = live[live.prglngth>30]  # filter the live df
resp = nsfg.ReadFemResp()
resp.index = resp.caseid
join = live.join(resp, on='caseid', rsuffix='_r')
join.shape  # return the shape of the df after the join
```

[8]:
```
(8884, 3331)
```

[11]:
```
# replace the value 97 in the df join with NaN, and compute the squared value of the age_r

join.numbabes.replace([97], np.nan, inplace=True)
join['age2'] = join.age_r**2
```

[12]:
```
# perform a Poisson regression analysis to model the count variable

formula='numbabes ~ age_r + age2 + age3 + C(race) + totincr + educat'
formula='numbabes ~ age_r + age2 + C(race) + totincr + educat'
model = smf.poisson(formula, data=join)
results = model.fit()
results.summary()
```

```
Optimization terminated successfully.
         Current function value: 1.677002
         Iterations 7
```

[12]:

Dep. Variable:	numbabes	**No. Observations:**	8884
Model:	Poisson	**Df Residuals:**	8877
Method:	MLE	**Df Model:**	6
Date:	Sat, 10 Feb 2024	**Pseudo R-squ.:**	0.03686
Time:	16:10:48	**Log-Likelihood:**	-14898.
converged:	True	**LL-Null:**	-15469.
Covariance Type:	nonrobust	**LLR p-value:**	3.681e-243

	coef	std err	z	P> |z|	[0.025	0.975]
Intercept	-1.0324	0.169	-6.098	0.000	-1.364	-0.701
C(race)[T.2]	-0.1401	0.015	-9.479	0.000	-0.169	-0.111
C(race)[T.3]	-0.0991	0.025	-4.029	0.000	-0.147	-0.051
age_r	0.1556	0.010	15.006	0.000	0.135	0.176
age2	-0.0020	0.000	-13.102	0.000	-0.002	-0.002
totincr	-0.0187	0.002	-9.830	0.000	-0.022	-0.015
educat	-0.0471	0.003	-16.076	0.000	-0.053	-0.041

[17]:
```
# predicted probabilities for each possible outcome of the dependent variable

columns = ['age_r', 'age2', 'age3', 'race', 'totincr', 'educat']
new = pd.DataFrame([[35, 35**2, 35**3, 1, 14, 16]], columns=columns)
results.predict(new)
```

[17]:

	0	1	2	3	4	5
0	0.782389	0.048214	0.001278	0.065284	0.032845	0.069991

11-4

[13]:
```
# perform a multinomial logistic regression analysis using the statsmodels
 ↪(SMF) library

formula='rmarital ~ age_r + age2 + C(race) + totincr + educat'
model = smf.mnlogit(formula, data=join)
results = model.fit()
results.summary()
```

```
Optimization terminated successfully.
         Current function value: 1.084053
         Iterations 8
```

[13]:

Dep. Variable:	rmarital	**No. Observations:**	8884
Model:	MNLogit	**Df Residuals:**	8849
Method:	MLE	**Df Model:**	30
Date:	Sat, 10 Feb 2024	**Pseudo R-squ.:**	0.1682
Time:	16:10:53	**Log-Likelihood:**	-9630.7
converged:	True	**LL-Null:**	-11579.
Covariance Type:	nonrobust	**LLR p-value:**	0.000

rmarital=2	coef	std err	z	P> |z|	[0.025	0.975]
Intercept	9.0156	0.805	11.199	0.000	7.438	10.593
C(race)[T.2]	-0.9237	0.089	-10.418	0.000	-1.097	-0.750
C(race)[T.3]	-0.6179	0.136	-4.536	0.000	-0.885	-0.351
age_r	-0.3635	0.051	-7.150	0.000	-0.463	-0.264
age2	0.0048	0.001	6.103	0.000	0.003	0.006
totincr	-0.1310	0.012	-11.337	0.000	-0.154	-0.108
educat	-0.1953	0.019	-10.424	0.000	-0.232	-0.159

rmarital=3	coef	std err	z	P> |z|	[0.025	0.975]
Intercept	2.9570	3.020	0.979	0.328	-2.963	8.877
C(race)[T.2]	-0.4411	0.237	-1.863	0.062	-0.905	0.023
C(race)[T.3]	0.0591	0.336	0.176	0.860	-0.600	0.718
age_r	-0.3177	0.177	-1.798	0.072	-0.664	0.029
age2	0.0064	0.003	2.528	0.011	0.001	0.011
totincr	-0.3258	0.032	-10.175	0.000	-0.389	-0.263
educat	-0.0991	0.048	-2.050	0.040	-0.194	-0.004

rmarital=4	coef	std err	z	P> |z|	[0.025	0.975]
Intercept	-3.5238	1.205	-2.924	0.003	-5.886	-1.162
C(race)[T.2]	-0.3213	0.093	-3.445	0.001	-0.504	-0.139
C(race)[T.3]	-0.7706	0.171	-4.509	0.000	-1.106	-0.436
age_r	0.1155	0.071	1.626	0.104	-0.024	0.255
age2	-0.0007	0.001	-0.701	0.483	-0.003	0.001
totincr	-0.2276	0.012	-19.621	0.000	-0.250	-0.205
educat	0.0667	0.017	3.995	0.000	0.034	0.099

rmarital=5	coef	std err	z	P> |z|	[0.025	0.975]
Intercept	-2.8963	1.305	-2.220	0.026	-5.453	-0.339
C(race)[T.2]	-1.0407	0.104	-10.038	0.000	-1.244	-0.837
C(race)[T.3]	-0.5661	0.156	-3.635	0.000	-0.871	-0.261
age_r	0.2411	0.079	3.038	0.002	0.086	0.397
age2	-0.0035	0.001	-2.977	0.003	-0.006	-0.001
totincr	-0.2932	0.015	-20.159	0.000	-0.322	-0.265
educat	-0.0174	0.021	-0.813	0.416	-0.059	0.025

rmarital=6	coef	std err	z	P> |z|	[0.025	0.975]
Intercept	8.0533	0.814	9.890	0.000	6.457	9.649
C(race)[T.2]	-2.1871	0.080	-27.211	0.000	-2.345	-2.030
C(race)[T.3]	-1.9611	0.138	-14.188	0.000	-2.232	-1.690
age_r	-0.2127	0.052	-4.122	0.000	-0.314	-0.112
age2	0.0019	0.001	2.321	0.020	0.000	0.003
totincr	-0.2945	0.012	-25.320	0.000	-0.317	-0.272
educat	-0.0742	0.018	-4.169	0.000	-0.109	-0.039

[14]:
```
# predict the probabilities for each possible outcome of the dependent variable

# This individual has a 75.0028% likelihood of being presently married, and
# a 12.6397% probability of residing with an opposite-sex partner without␣
 ↪marriage

columns = ['age_r', 'age2', 'race', 'totincr', 'educat']
new = pd.DataFrame([[25, 25**2, 2, 11, 12]], columns=columns)
results.predict(new)
```

[14]:

	0	1	2	3	4	5
0	0.750028	0.126397	0.001564	0.033403	0.021485	0.067122

[ ]: