In [1]:
```
# DSC530-T302
# Stephen Smitshoek
# Week08
# Exercise 11-1
```

In [2]:
```
import thinkstats2
import pandas
import numpy as np
import statsmodels.formula.api as smf
import first
import nsfg
```

In [3]:
```
live, firsts, others = first.MakeFrames()
live = live[live.prglngth>30]

resp = nsfg.ReadFemResp()
resp.index = resp.caseid

join_live_resp = live.join(resp, on='caseid', rsuffix='_r')
```

In [4]:
```
model = smf.ols('prglngth ~ nbrnaliv>1', data=join_live_resp)
results = model.fit()
print(results.summary()) # R2 = 0.009 P = 0.000
```

```
                            OLS Regression Results
==============================================================================
Dep. Variable:               prglngth   R-squared:                       0.009
Model:                            OLS   Adj. R-squared:                  0.009
Method:                 Least Squares   F-statistic:                     84.63
Date:                Wed, 27 Jul 2022   Prob (F-statistic):           4.41e-20
Time:                        07:00:08   Log-Likelihood:                -18256.
No. Observations:                8884   AIC:                         3.652e+04
Df Residuals:                    8882   BIC:                         3.653e+04
Df Model:                           1
Covariance Type:            nonrobust
===============================================================================
                  coef    std err          t      P>|t|      [0.025      0.975]
-------------------------------------------------------------------------------
Intercept      38.9009      0.020   1926.120      0.000      38.861      38.941
nbrnaliv > 1[T.True]    -1.5129      0.164     -9.200      0.000      -1.835      -1.190
==============================================================================
Omnibus:                     1556.232   Durbin-Watson:                   1.621
Prob(Omnibus):                  0.000   Jarque-Bera (JB):             6037.784
Skew:                          -0.835   Prob(JB):                         0.00
Kurtosis:                       6.678   Cond. No.                         8.21
==============================================================================

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

In [5]:
```
model = smf.ols('prglngth ~ C(race)', data=join_live_resp)
```

```
results = model.fit()
print(results.summary()) # R2 = 0.001 P = 0.001 (For C=2)
```

```
                            OLS Regression Results
==============================================================================
Dep. Variable:                prglngth   R-squared:                       0.001
Model:                            OLS   Adj. R-squared:                  0.001
Method:                 Least Squares   F-statistic:                     5.752
Date:                Wed, 27 Jul 2022   Prob (F-statistic):            0.00319
Time:                        07:00:08   Log-Likelihood:                -18293.
No. Observations:                8884   AIC:                         3.659e+04
Df Residuals:                    8881   BIC:                         3.661e+04
Df Model:                           2
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
Intercept      38.7798      0.039    986.735      0.000      38.703      38.857
C(race)[T.2]    0.1483      0.047      3.185      0.001       0.057       0.240
C(race)[T.3]    0.0232      0.078      0.297      0.766      -0.130       0.176
==============================================================================
Omnibus:                     1599.505   Durbin-Watson:                   1.631
Prob(Omnibus):                  0.000   Jarque-Bera (JB):             6170.529
Skew:                          -0.860   Prob(JB):                         0.00
Kurtosis:                       6.703   Cond. No.                         5.20
==============================================================================

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

In [6]:
```
model = smf.ols('prglngth ~ agepreg', data=join_live_resp)
results = model.fit()
print(results.summary()) # R2 = 0.000 P = 0.693
```

```
                            OLS Regression Results
==============================================================================
Dep. Variable:               prglngth   R-squared:                       0.000
Model:                            OLS   Adj. R-squared:                 -0.000
Method:                 Least Squares   F-statistic:                    0.1564
Date:                Wed, 27 Jul 2022   Prob (F-statistic):              0.693
Time:                        07:00:08   Log-Likelihood:                -18299.
No. Observations:                8884   AIC:                         3.660e+04
Df Residuals:                    8882   BIC:                         3.662e+04
Df Model:                           1
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
Intercept     38.9138      0.093    420.444      0.000      38.732      39.095
agepreg       -0.0014      0.004     -0.395      0.693      -0.009       0.006
==============================================================================
Omnibus:                     1586.727   Durbin-Watson:                   1.630
Prob(Omnibus):                  0.000   Jarque-Bera (JB):             6125.870
Skew:                          -0.853   Prob(JB):                         0.00
Kurtosis:                       6.693   Cond. No.                         118.
==============================================================================

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

In [7]:
```
model = smf.ols('prglngth ~ birthord', data=join_live_resp)
results = model.fit()
print(results.summary()) # R2 = 0.001 P = 0.001
```

```
                            OLS Regression Results
==============================================================================
Dep. Variable:               prglngth   R-squared:                       0.001
Model:                            OLS   Adj. R-squared:                  0.001
Method:                 Least Squares   F-statistic:                     11.00
Date:                Wed, 27 Jul 2022   Prob (F-statistic):           0.000913
Time:                        07:00:08   Log-Likelihood:                -18293.
No. Observations:                8884   AIC:                         3.659e+04
Df Residuals:                    8882   BIC:                         3.660e+04
Df Model:                           1
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
Intercept     38.9953      0.041    958.709      0.000      38.916      39.075
birthord      -0.0642      0.019     -3.317      0.001      -0.102      -0.026
==============================================================================
Omnibus:                     1607.192   Durbin-Watson:                   1.628
Prob(Omnibus):                  0.000   Jarque-Bera (JB):             6188.998
Skew:                          -0.865   Prob(JB):                         0.00
Kurtosis:                       6.705   Cond. No.                         5.01
==============================================================================

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

In [8]:
```
model = smf.ols('prglngth ~ babysex==1', data=join_live_resp)
results = model.fit()
```

```
print(results.summary()) # R2 = 0.000 P = 0.187
```

```
                            OLS Regression Results
==============================================================================
Dep. Variable:               prglngth   R-squared:                       0.000
Model:                            OLS   Adj. R-squared:                  0.000
Method:                 Least Squares   F-statistic:                     1.745
Date:                Wed, 27 Jul 2022   Prob (F-statistic):              0.187
Time:                        07:00:08   Log-Likelihood:                -18298.
No. Observations:                8884   AIC:                         3.660e+04
Df Residuals:                    8882   BIC:                         3.661e+04
Df Model:                           1
Covariance Type:            nonrobust
================================================================================================
                                   coef    std err          t      P>|t|      [0.025      0.9
75]
------------------------------------------------------------------------------------------------
---
Intercept                       38.8511      0.029   1353.307      0.000      38.795      38.
907
babysex == 1[T.True]             0.0532      0.040      1.321      0.187      -0.026       0.
132
==============================================================================
Omnibus:                     1584.119   Durbin-Watson:                   1.631
Prob(Omnibus):                  0.000   Jarque-Bera (JB):             6103.809
Skew:                          -0.852   Prob(JB):                         0.00
Kurtosis:                       6.686   Cond. No.                         2.64
==============================================================================

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

In [9]:

```
model = smf.ols('prglngth ~ paydu==1', data=join_live_resp)
results = model.fit()
print(results.summary()) # R2 = 0.001 P = 0.003
```

```
                            OLS Regression Results
==============================================================================
Dep. Variable:               prglngth   R-squared:                       0.001
Model:                            OLS   Adj. R-squared:                  0.001
Method:                 Least Squares   F-statistic:                     8.917
Date:                Wed, 27 Jul 2022   Prob (F-statistic):            0.00283
Time:                        07:00:08   Log-Likelihood:                -18294.
No. Observations:                8884   AIC:                         3.659e+04
Df Residuals:                    8882   BIC:                         3.661e+04
Df Model:                           1
Covariance Type:            nonrobust
===============================================================================================
                                  coef    std err          t      P>|t|      [0.025      0.97
5]
-----------------------------------------------------------------------------------------------
-
Intercept                    38.8189      0.028   1374.378      0.000      38.764      38.87
4
paydu == 1[T.True]            0.1202      0.040      2.986      0.003       0.041       0.19
9
==============================================================================
Omnibus:                     1586.238   Durbin-Watson:                   1.630
Prob(Omnibus):                  0.000   Jarque-Bera (JB):             6129.329
Skew:                          -0.852   Prob(JB):                         0.00
Kurtosis:                       6.695   Cond. No.                         2.60
==============================================================================

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

In [10]:

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

Out[10]:

OLS Regression Results

Dep. Variable:	prglngth	**R-squared:**	0.012
Model:	OLS	**Adj. R-squared:**	0.011
Method:	Least Squares	**F-statistic:**	26.72
Date:	Wed, 27 Jul 2022	**Prob (F-statistic):**	4.54e-22
Time:	07:00:08	**Log-Likelihood:**	-18245.
No. Observations:	8884	**AIC:**	3.650e+04
Df Residuals:	8879	**BIC:**	3.654e+04
Df Model:	4		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
Intercept	38.8749	0.053	733.490	0.000	38.771	38.979
nbrnaliv > 1[T.True]	-1.4908	0.164	-9.063	0.000	-1.813	-1.168
paydu == 1[T.True]	0.0821	0.042	1.974	0.048	0.001	0.164
race == 2[T.True]	0.1107	0.044	2.539	0.011	0.025	0.196
birthord	-0.0474	0.019	-2.442	0.015	-0.085	-0.009

Omnibus:	1583.093	**Durbin-Watson:**	1.619
Prob(Omnibus):	0.000	**Jarque-Bera (JB):**	6132.823
Skew:	-0.850	**Prob(JB):**	0.00
Kurtosis:	6.699	**Cond. No.**	19.7

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