In [1]:

```python
import numpy as np
import pandas as pd
import statsmodels.api as sm
```

In [2]:

```python
pd.set_option('display.max_rows',None)
pd.set_option('display.max_columns',None)
```

In [5]:

```python
df = pd.read_excel('05-19data/制造业工业表.xlsx')
```

In [6]:

```python
df
```

Out[6]:

	year	出荷額	中間投入	資本サービス投入指数	実質資本ストック指数
0	2009	233860164	134270200	1.031655	1.027320
1	2010	258645824	148766344	1.009572	1.008295
2	2011	252549212	147308571	1.000000	1.000000
3	2012	257281277	149299186	0.995938	0.995815
4	2013	260826123	151820385	0.988484	0.990102
5	2014	272904301	161253005	0.991805	0.992691
6	2015	279553100	159511845	1.005195	1.003080
7	2016	270703974	149408500	1.018142	1.014045
8	2017	287074969	161237634	1.031694	1.025467
9	2018	299705734	171506083	1.052968	1.042834

In [7]:

```python
df2 = pd.read_excel('05-19data/washed_data.xlsx',index_col=0)
```

In [11]:

```python
data = pd.merge(left = df,right = df2,on = 'year',how = 'outer')
```

In [18]:

```python
data.drop(columns = '学历年龄',inplace=True)
```

In [21]:

```python
len(data.columns)
```

Out[21]:

```
15
```

In [22]:

```python
data['ckb'] = data['資本サービス投入指数']*data['実質資本ストック指数']
```

In [23]:

```python
data['total_working_time'] = data['所定内実労働時間数'] + data['超過実労働時間数']
```

In [24]:

```python
data['salary'] = data['きまって支給する現金給与額(千円)'] *12 +data['年間賞与その他特別給与額(千円)']
```

In [25]:

```python
data['勤続年数2乗'] = np.square(data['勤続年数'])
```

In [26]:

```python
data = data.astype(np.float64)
```

In [28]:

```python
data['実際中間投入'] = np.log(data['中间投入'])
```

In [34]:

```python
data['実質産出'] = np.log(data['出荷額'])
```

In [45]:

```python
 male_data = data[data['sex']==1]
```

In [47]:

```python
male_data.reset_index(drop = 'index',inplace=True)
```

In [54]:

```python
female_data = data[data['sex'] ==0]
```

In [55]:

```python
female_data.reset_index(inplace=True,drop = 'index')
```

In [56]:

```python
female_data
```

4	2009.0	233860164.0	134270200.0	1.031655	1.027320	37.5	10.0	166.0	7.0	174.6	165.8	334.9	298.0	9.0	0.0	1.059840	173.0	2430.1
5	2009.0	233860164.0	134270200.0	1.031655	1.027320	42.7	9.6	162.0	8.0	175.7	166.6	270.4	262.0	9.0	0.0	1.059840	170.0	2378.8
6	2009.0	233860164.0	134270200.0	1.031655	1.027320	47.6	15.9	168.0	5.0	202.0	192.3	431.1	346.0	9.0	0.0	1.059840	173.0	2855.1
7	2009.0	233860164.0	134270200.0	1.031655	1.027320	52.8	21.1	164.0	5.0	206.8	199.0	620.6	900.0	9.0	0.0	1.059840	169.0	3102.2
8	2009.0	233860164.0	134270200.0	1.031655	1.027320	57.6	23.2	163.0	5.0	209.6	201.4	576.1	2289.0	9.0	0.0	1.059840	168.0	3091.3
9	2009.0	233860164.0	134270200.0	1.031655	1.027320	41.3	12.8	164.0	7.0	209.2	198.6	573.0	61080.0	12.0	0.0	1.059840	171.0	3083.4
10	2009.0	233860164.0	134270200.0	1.031655	1.027320	22.3	3.4	167.0	10.0	181.3	167.8	464.0	5521.0	12.0	0.0	1.059840	177.0	2639.6
11	2009.0	233860164.0	134270200.0	1.031655	1.027320	27.6	6.5	165.0	8.0	196.0	184.3	550.7	5614.0	12.0	0.0	1.059840	173.0	2902.7
12	2009.0	233860164.0	134270200.0	1.031655	1.027320	32.5	10.0	163.0	7.0	207.4	196.4	601.2	6162.0	12.0	0.0	1.059840	170.0	3090.0
13	2009.0	233860164.0	134270200.0	1.031655	1.027320	37.4	12.8	162.0	6.0	222.5	212.1	687.0	8295.0	12.0	0.0	1.059840	168.0	3357.0
14	2009.0	233860164.0	134270200.0	1.031655	1.027320	42.5	13.7	162.0	7.0	223.8	211.5	664.6	8211.0	12.0	0.0	1.059840	169.0	3350.2
15	2009.0	233860164.0	134270200.0	1.031655	1.027320	47.6	15.7	164.0	6.0	223.4	213.8	669.8	7516.0	12.0	0.0	1.059840	170.0	3342.?

In [ ]:

女性モデル

```
log(y(t)) = C +  + α1*exp + α2*exp_square + α3*edu + α4*log(ME(t)) + α5*log(CKB(t)) + α6*Full  + u
y: 実質産出
exp:勤続年数
exp_square:勤続年数2乗
edu:教育年数
ME(t): 実際中間投入(t年での条件)
CKB(t): 資本サービス投入指数 * 実質資本ストック指数(t年での条件)
Full: 総労動时间
```

In [69]:

```python
female_data.head()
```

Out[69]:

	year	出荷額	中間投入	資本サービス投入指数	実質資本ストック指数	年齢	勤続年数	所定内実労働時間数	超過実労働時間数	きまって支給する現金給与額(千円)	所定内給与額(千円)	年間賞与その他特別給与額(千円)	労働者数(十人)	education	sex	ckb	total_working_time	s
0	2009.0	233860164.0	134270200.0	1.031655	1.02732	51.9	17.2	166.0	7.0	186.7	177.8	408.8	6639.0	9.0	0.0	1.05984	173.0	2
1	2009.0	233860164.0	134270200.0	1.031655	1.02732	22.5	1.6	173.0	20.0	146.6	127.0	40.9	318.0	9.0	0.0	1.05984	193.0	1
2	2009.0	233860164.0	134270200.0	1.031655	1.02732	27.7	2.6	176.0	8.0	139.7	132.8	80.1	309.0	9.0	0.0	1.05984	184.0	1
3	2009.0	233860164.0	134270200.0	1.031655	1.02732	32.8	5.4	167.0	8.0	150.5	142.5	154.3	304.0	9.0	0.0	1.05984	175.0	1
4	2009.0	233860164.0	134270200.0	1.031655	1.02732	37.5	10.0	166.0	7.0	174.6	165.8	334.9	298.0	9.0	0.0	1.05984	173.0	2

In [79]:

```python
y = female_data['実質産出']
```

In [80]:

```python
feature  = female_data.loc[:,['勤続年数','勤続年数2乗','education','実際中間投入','ckb','total_working_time']]
```

In [81]:

```python
feature = sm.add_constant(feature)
```

In [82]:

```python
res = sm.OLS(y,feature).fit()
```

In [83]:

```
res.summary()
```

Out[83]:

OLS Regression Results

Dep. Variable:	実質産出	**R-squared:**	0.947
Model:	OLS	**Adj. R-squared:**	0.946
Method:	Least Squares	**F-statistic:**	1043.
Date:	Wed, 01 Feb 2023	**Prob (F-statistic):**	3.36e-221
Time:	15:32:38	**Log-Likelihood:**	989.45
No. Observations:	360	**AIC:**	-1965.
Df Residuals:	353	**BIC:**	-1938.
Df Model:	6		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
const	0.8552	0.254	3.369	0.001	0.356	1.355
勤続年数	0.0004	0.001	0.761	0.447	-0.001	0.002
勤続年数2乗	-8.81e-06	2.01e-05	-0.439	0.661	-4.83e-05	3.07e-05
education	0.0005	0.000	1.215	0.225	-0.000	0.001
実際中間投入	0.9640	0.015	66.458	0.000	0.936	0.993
ckb	0.2934	0.023	12.495	0.000	0.247	0.340
total_working_time	0.0004	0.000	1.806	0.072	-3.38e-05	0.001

Omnibus:	48.381	**Durbin-Watson:**	0.050
Prob(Omnibus):	0.000	**Jarque-Bera (JB):**	64.803
Skew:	1.029	**Prob(JB):**	8.47e-15
Kurtosis:	3.287	**Cond. No.**	9.08e+04

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
[2] The condition number is large, 9.08e+04. This might indicate that there are
strong multicollinearity or other numerical problems.

男性モデル

```
log(y(t)) = C +  + α1*exp + α2*exp_square + α3*edu + α4*log(ME(t)) + α5*log(CKB(t)) + α6*Full  + u
y: 実質産出
exp:勤続年数
exp_square:勤続年数2乗
edu:教育年数
ME(t): 実際中間投入(t年での条件)
CKB(t): 資本サービス投入指数  * 実質資本ストック指数(t年での条件)
Full: 総労動时间
```

In [84]:

```
male_data.head()
```

Out[84]:

	year	出荷額	中間投入	資本サービス投入指数	実質資本ストック指数	年齢	勤続年数	所定内実労働時間数	超過実労働時間数	きまって支給する現金給与額(千円)	所定内給与額(千円)	年間賞与その他特別給与額(千円)	労働者数(十人)	education	sex	ckb	total_working_time
0	2009.0	233860164.0	134270200.0	1.031655	1.02732	50.1	22.1	164.0	12.0	318.0	290.8	920.1	22719.0	9.0	1.0	1.05984	176.0
1	2009.0	233860164.0	134270200.0	1.031655	1.02732	22.7	2.5	168.0	34.0	222.0	180.8	313.8	852.0	9.0	1.0	1.05984	202.0
2	2009.0	233860164.0	134270200.0	1.031655	1.02732	27.5	5.3	165.0	16.0	241.6	212.1	543.2	1081.0	9.0	1.0	1.05984	181.0
3	2009.0	233860164.0	134270200.0	1.031655	1.02732	32.7	8.3	167.0	16.0	263.5	232.3	646.3	1208.0	9.0	1.0	1.05984	183.0
4	2009.0	233860164.0	134270200.0	1.031655	1.02732	37.3	11.6	166.0	16.0	295.9	261.1	753.3	1814.0	9.0	1.0	1.05984	182.0

In [85]:

```
y = male_data['実質産出']
```

In [86]:

```
feature  = female_data.loc[:,['勤続年数','勤続年数2乗','education','実際中間投入','ckb','total_working_time']]
```

In [87]:

```
feature = sm.add_constant(feature)
```

In [88]:

```
res = sm.OLS(y,feature).fit()
```

In [89]:

```
res.summary()
```

Out[89]:

OLS Regression Results

Dep. Variable:	実質産出	R-squared:	0.947
Model:	OLS	Adj. R-squared:	0.946
Method:	Least Squares	F-statistic:	1043.
Date:	Wed, 01 Feb 2023	Prob (F-statistic):	3.36e-221
Time:	15:34:45	Log-Likelihood:	989.45
No. Observations:	360	AIC:	-1965.
Df Residuals:	353	BIC:	-1938.
Df Model:	6		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
const	0.8552	0.254	3.369	0.001	0.356	1.355
勤続年数	0.0004	0.001	0.761	0.447	-0.001	0.002
勤続年数2乗	-8.81e-06	2.01e-05	-0.439	0.661	-4.83e-05	3.07e-05
education	0.0005	0.000	1.215	0.225	-0.000	0.001
実際中間投入	0.9640	0.015	66.458	0.000	0.936	0.993
ckb	0.2934	0.023	12.495	0.000	0.247	0.340
total_working_time	0.0004	0.000	1.806	0.072	-3.38e-05	0.001

Omnibus:	48.381	Durbin-Watson:	0.050
Prob(Omnibus):	0.000	Jarque-Bera (JB):	64.803
Skew:	1.029	Prob(JB):	8.47e-15
Kurtosis:	3.287	Cond. No.	9.08e+04

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
[2] The condition number is large, 9.08e+04. This might indicate that there are
strong multicollinearity or other numerical problems.

In [ ]: