The effect of Gross Domestic Product on food imports

In [2]:
```
#Working directory
os.chdir('/Users/duccioa/CLOUD/C07_UCL_SmartCities/QuantitativeMethods/qm_cours
ewrk1')
#Load packages
import matplotlib.pyplot as plt
import matplotlib
matplotlib.style.use('ggplot')
import patsy
import pandas as pd
#import statsmodels as sm
#import statsmodels.api as sm
import numpy as np
import statsmodels.api as sm
from statsmodels.formula.api import ols
import csv
import os
#Set defaults for graphs
pd.set_option('display.mpl_style', 'default') # Make the graphs a bit prettier
figsize(10, 5)
```

Introduction

In this project, I investigate the determinants of food imports in countries based on the dataset Data_for_Coursework_1_Countries.csv, provided within the context of the course Quantitative Methods at UCL. A first glance at the data reveals an interesting sublinear power law relation between GDP and food imports. Countries with bigger economies import more food then countries with a smaller output but the increase is less important the more the GDP grows. At a second glance, the same relation holds if we take into account population but with an effect of economy of scale due to the population: richer countries import more than poorer ones but spend a smaller fraction of their income in food, bigger countries import less food per capita than smaller ones.

Analysis

In [4]:
```
countries = pd.read_csv('countries.csv', sep = ',', encoding = 'latin1')
countries.columns = ['X', 'Year', 'CountryCode', 'CountryName',
                     'Population', 'GDP', 'FoodImports', 'FuelImports']
```

In this project I aim at investigating the determinant of food import based on the available dataset, which contains data for 190 countries about Gross Domestic Product (GDP), population, food and fuel imports for the year 2005.

In [5]:
```
countries[:3]
```

Out[5]:

	X	Year	CountryCode	CountryName	Population	GDP	FoodImports	FuelImports
0	1	2005	ABW	Aruba	100031	1.160240e+12	97166150	32335285
1	2	2005	AFG	Afghanistan	24860855	6.275076e+09	528341972	461521897
2	3	2005	AGO	Angola	16544376	2.823370e+10	1075607744	48218538

In [6]:
```
countries['Population_log'] = log(countries['Population'])
countries['FoodImports_log'] = log(countries['FoodImports'])
countries['GDP_log'] = log(countries['GDP'])
countries['GDP_pc'] = countries['GDP']/countries['Population']
countries['FoodImports_pc'] = countries['FoodImports']/countries['Population']
countries['GDP_pc_log'] = log(countries['GDP_pc'])
countries['FoodImports_pc_log'] = log(countries['FoodImports_pc'])
```

Both plots present patterns of correlation with three big outliers: USA, India and China which have respectivly very high GDP, population, GDP and population. GDP have a very strong correlation (0.92) with food import and the ln-ln plot strongly suggests a power law relation with values very concentrated on the regression line (the ln-ln plot with population shows higher variation than with GDP).

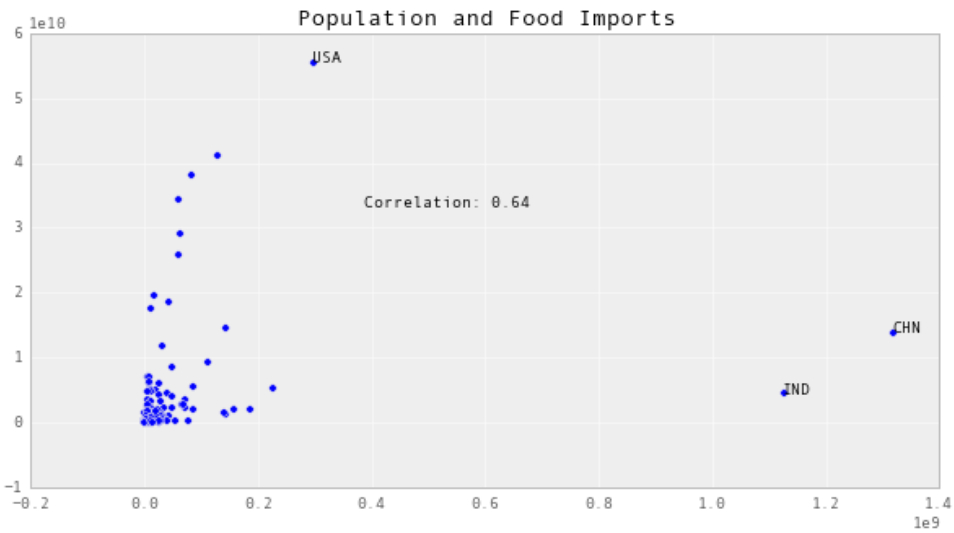
In [7]:
```
coords = countries[countries['CountryCode'] == 'CHN']
coords = coords.append(countries[countries['CountryCode'] == 'IND'])
coords = coords.append(countries[countries['CountryCode'] == 'USA'])
```

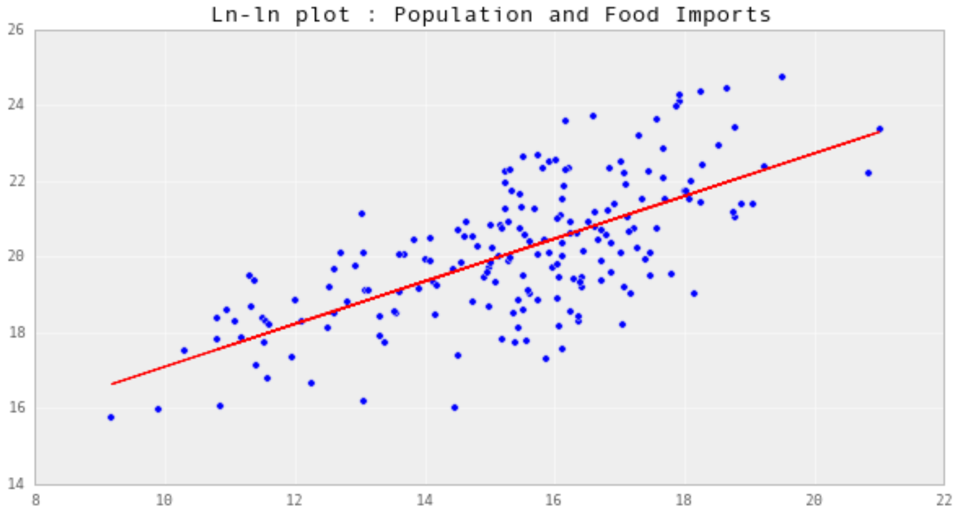
In [8]:

```
#POPULATION AND FOOD IMPORT : analysis and plot
fig, ax = plt.subplots()
ax.scatter(countries['Population'], countries['FoodImports'])
ax.annotate(coords['CountryCode'][179],
            xy = (coords['Population'][179], coords['FoodImports'][179]))
ax.annotate(coords['CountryCode'][34],
            xy = (coords['Population'][34], coords['FoodImports'][34]))
ax.annotate(coords['CountryCode'][82],
            xy = (coords['Population'][82], coords['FoodImports'][82]))
ax.annotate('Correlation: ' +
            str(round(countries['Population'].corr(countries['FoodImports'], me
thod='spearman'), 2)),
            xy = (coords['Population'][179]*1.3, coords['FoodImports'][179]*0.6
))
plt.title('Population and Food Imports')
#Log plot
fig, ax = plt.subplots()
ax.scatter(countries['Population_log'], countries['FoodImports_log'])
plt.title('Ln-ln plot : Population and Food Imports')

#Regression line
mod_pop = ols(formula='FoodImports_log ~ Population_log', data=countries)
res_pop = mod_pop.fit()
par_pop = res_pop.params
ax.plot(countries['Population_log'], par_pop[0] +
        par_pop[1]*countries['Population_log'], color = 'red')
```

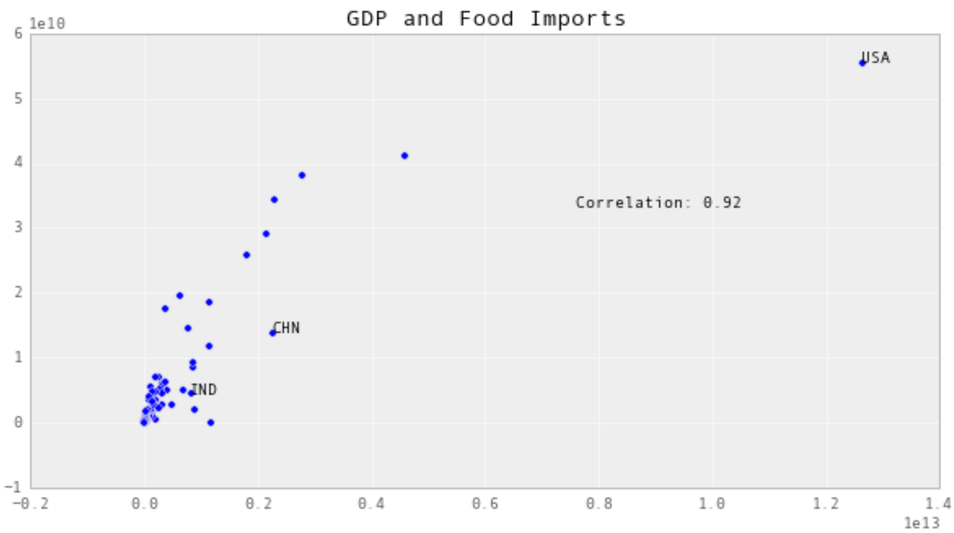
```
Out[8]: [<matplotlib.lines.Line2D at 0x110c2de50>]
```

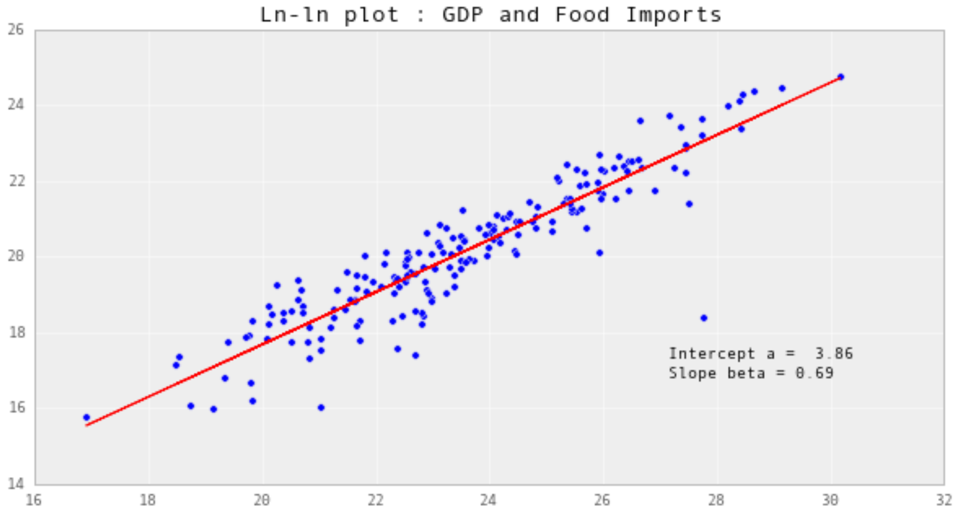




In [9]:
```
##GDP AND FOOD IMPORT : analysis and plot
#Plot
fig, ax = plt.subplots()
ax.scatter(countries['GDP'], countries['FoodImports'])
ax.annotate(coords['CountryCode'][179],
            xy = (coords['GDP'][179], coords['FoodImports'][179]))
ax.annotate(coords['CountryCode'][34],
            xy = (coords['GDP'][34], coords['FoodImports'][34]))
ax.annotate(coords['CountryCode'][82],
            xy = (coords['GDP'][82], coords['FoodImports'][82]))
ax.annotate('Correlation: ' +
            str(round(countries['GDP'].corr(countries['FoodImports'],
                                            method='spearman'), 2)),
            xy = (coords['GDP'][179]*0.6, coords['FoodImports'][179]*0.6))
plt.title('GDP and Food Imports')
#Log plot
fig, ax = plt.subplots()
ax.scatter(countries['GDP_log'], countries['FoodImports_log'])
plt.title('Ln-ln plot : GDP and Food Imports')
#Regression line
mod_gdp = ols(formula='FoodImports_log ~ GDP_log', data=countries)
res_gdp = mod_gdp.fit()
par_gdp = res_gdp.params
ax.plot(countries['GDP_log'], par_gdp[0] +
        par_gdp[1]*countries['GDP_log'], color = 'red')
ax.annotate('Intercept a =  ' + str(round(par_gdp[0], 2)),
            xy = (coords['GDP_log'][179]*0.9, coords['FoodImports_log'][179]*0.
7))
ax.annotate('Slope beta = '  + str(round(par_gdp[1], 2)),
            xy = (coords['GDP_log'][179]*0.9, coords['FoodImports_log'][179]*0.
68))
```

```
Out[9]: <matplotlib.text.Annotation at 0x110fbb050>
```





Based on the graphs above the relation between the outcome Foor Imports (Y) and the predictor GDP (X) is as following:

$$Y = aX^{\beta}$$

In [10]:
```
print res_gdp.summary()
```

```
                            OLS Regression Results
==============================================================================
Dep. Variable:        FoodImports_log   R-squared:                       0.842
Model:                            OLS   Adj. R-squared:                  0.841
Method:                 Least Squares   F-statistic:                     1001.
Date:                Sun, 01 Nov 2015   Prob (F-statistic):           3.27e-77
Time:                        17:59:31   Log-Likelihood:                -208.20
No. Observations:                 190   AIC:                             420.4
Df Residuals:                     188   BIC:                             426.9
Df Model:                           1
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [95.0% Conf. Int.]
------------------------------------------------------------------------------
Intercept      3.8598      0.517      7.463      0.000         2.839     4.880
GDP_log        0.6912      0.022     31.634      0.000         0.648     0.734
==============================================================================
Omnibus:                      101.220   Durbin-Watson:                   1.832
Prob(Omnibus):                  0.000   Jarque-Bera (JB):              695.506
Skew:                          -1.905   Prob(JB):                    9.40e-152
Kurtosis:                      11.564   Cond. No.                         232.
==============================================================================

Warnings:
[1] Standard Errors assume that the covariance matrix of the errors is correct
ly specified.
```

With a = 3.86 and α = 0.69, the model explains almost 85% of the variation in food import (R-squared = 0.84), which is a good fit. Being the slope < 1, the relation is sublinear and the impact of GDP on food import declines with the increase of GDP. However we would expect both GDP and food imports to be strongly increasing in the population. So it is worth to check that this strong correlation is not driven by their dependence on population. To do so I run the following regression:

$$ln(y) = \alpha_1 + \alpha_2 ln(x) + \alpha_3 ln(L)$$

With $y = \frac{Y}{L}$ (food import per capita) and $x = \frac{X}{L}$ (GDP per capita).
It is worth to note that this formulation encompasses the one above (for $\alpha_2 = 1 - \alpha_3 = \beta$ it is identical, see appendix for details).

In [11]:
```
mod = ols(formula='FoodImports_pc_log ~ GDP_pc_log + Population_log', data=coun
tries)
res = mod.fit()
print res.summary()
```

```
                             OLS Regression Results
==============================================================================
Dep. Variable:     FoodImports_pc_log   R-squared:                       0.807
Model:                            OLS   Adj. R-squared:                  0.805
Method:                 Least Squares   F-statistic:                     390.9
Date:                Sun, 01 Nov 2015   Prob (F-statistic):           1.60e-67
Time:                        17:59:34   Log-Likelihood:                -207.91
No. Observations:                 190   AIC:                             421.8
Df Residuals:                     187   BIC:                             431.6
Df Model:                           2
Covariance Type:            nonrobust
==============================================================================
====
                     coef    std err          t      P>|t|      [95.0% Conf. I
nt.]
------------------------------------------------------------------------------
----
Intercept          3.8607      0.518      7.456      0.000         2.839     4
.882
GDP_pc_log         0.6740      0.032     21.300      0.000         0.612     0
.736
Population_log    -0.2998      0.025    -12.010      0.000        -0.349    -0
.251
==============================================================================
Omnibus:                       92.805   Durbin-Watson:                   1.841
Prob(Omnibus):                  0.000   Jarque-Bera (JB):              534.208
Skew:                          -1.778   Prob(JB):                    9.96e-117
Kurtosis:                      10.405   Cond. No.                         172.
==============================================================================

Warnings:
[1] Standard Errors assume that the covariance matrix of the errors is correct
ly specified.
```

This gives us the following equation :

$$y = ax^{\alpha_2} L^{\alpha_3}$$

With $y = \frac{Y}{L}$ (food import per capita) and $x = \frac{X}{L}$ (GDP per capita), a = 47.49, $\alpha_2 = 0.67$ and $\alpha_3 = 0.3$.
The model has an R-squared = 0.807 and the coefficients are all statistically significant.

The equation above suggests that, holding population constant, the food import per capita increases in a sublinear fashion with the GDP per capita and the increase is less and less pronounced for more productive countries (higher GDP per capita). On the other hand, holding GDP per capita constant, countries with more population import less food per person. The first relation is probably explained by the fact that as people become richer, they consume more food but spend a smaller fraction of their income in buying it. The other relation could be explained as an economy of scale, where increase in population make food allocation more efficient and reduces the realtive need of importing food.
The model explains $\sim$80% of the variation and other factors would be worth investigating. To name a few, I would think food trade balance, levels of producivity and employment in agriculture would be a good start to improve the comprehension of the dynamics behind food imports.

Appendix

Identity of the power law

The power law suggested by the graph is

$$\left[Y = aX^{\beta}\right]$$

Dividing by the population L, we can write:

$$\frac{Y}{L} = \frac{aX^{\beta}}{L} = \frac{aX^{\beta}}{L^{\beta}L^{1-\beta}} = a(\frac{X}{L})^{\beta}L^{1-\beta}$$

If I write $y = \frac{Y}{L}$ as the Food Imports per capita and $x = \frac{X}{L}$ as the GDP per capita, I can say that the first equation is equivalent to saying:

$$\left[y = ax^{\beta}L^{1-\beta}\right]$$