In [1]:

```python
import numpy as np
import pandas as pd
import matplotlib as mpl
import matplotlib.pyplot as plt
import seaborn as sns
from sklearn.linear_model import LinearRegression
import statsmodels.formula.api as smf

%matplotlib inline
sns.set(style="white", font_scale=2.0)
```

Build a linear regression of predicting the body mass from these diameters.

In [2]:

```python
# read in data and get mass and other diameter columns
data = pd.read_csv('physical.txt',delimiter='\t')
x = data[['Fore','Bicep','Chest','Neck','Shoulder','Waist','Height','Calf',
'Thigh','Head']]
y = data['Mass']

data.head()
```

Out[2]:

	Mass	Fore	Bicep	Chest	Neck	Shoulder	Waist	Height	Calf	Thigh	Head
0	77.0	28.5	33.5	100.0	38.5	114.0	85.0	178.0	37.5	53.0	58.0
1	85.5	29.5	36.5	107.0	39.0	119.0	90.5	187.0	40.0	52.0	59.0
2	63.0	25.0	31.0	94.0	36.5	102.0	80.5	175.0	33.0	49.0	57.0
3	80.5	28.5	34.0	104.0	39.0	114.0	91.5	183.0	38.0	50.0	60.0
4	79.5	28.5	36.5	107.0	39.0	114.0	92.0	174.0	40.0	53.0	59.0

a) Plot the residual against the fitted values for your regression

In [3]:

```python
# make a linear regression model and fit with x and y
model = LinearRegression()
model.fit(x,y)

yf = model.predict(x)

residual = y-yf
```

In [4]:

```python
# plot values
ax = plt.subplot()
```
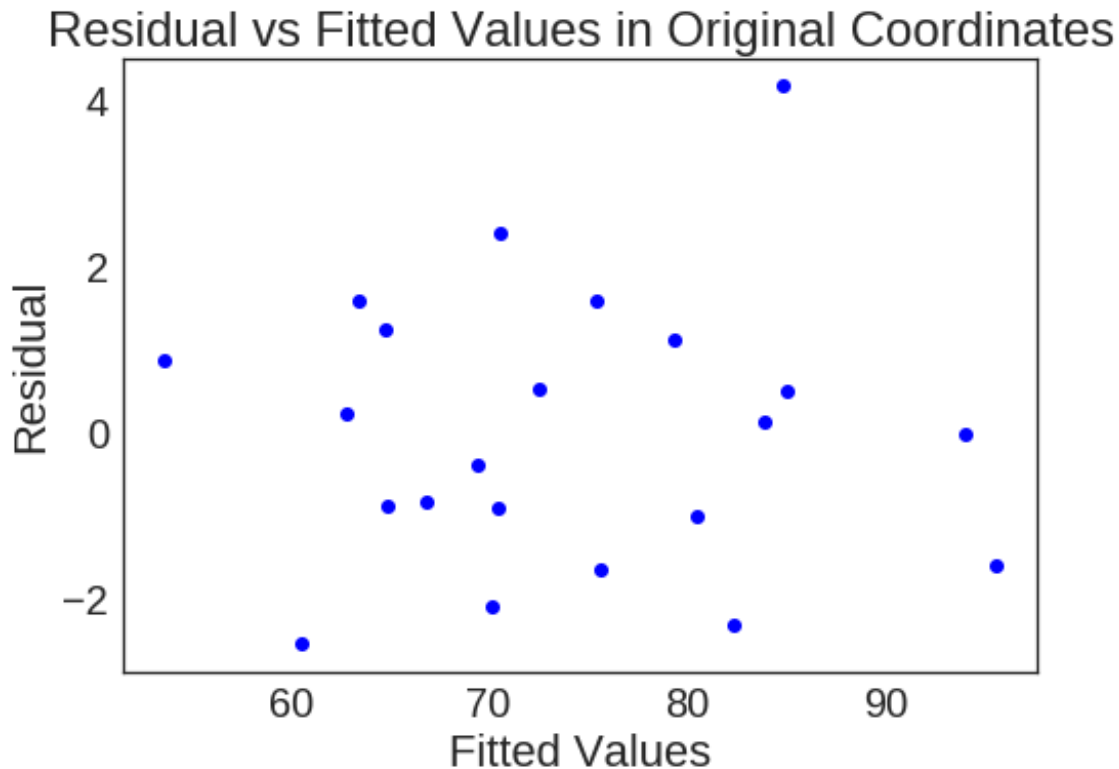
```
plt.plot(yf,residual,'o',color='b')

ax.set(xlabel='Fitted Values',
        ylabel='Residual',
        title='Residual vs Fitted Values in Original Coordinates')
```

Out[4]:

```
[<matplotlib.text.Text at 0x7fd09471b358>,
 <matplotlib.text.Text at 0x7fd0946fcb70>,
 <matplotlib.text.Text at 0x7fd0946c5320>]
```


Residual vs Fitted Values in Original Coordinates

b) Now regress the cube root of mass against these diameters. Plot the residual against the fitted values in both these cube root coordinates and in the original coordinates.

In [5]:

```
# get a new model for the cube rooted mass
model2 = LinearRegression()
model2.fit(x,y**(1/3))

yf2 = model.predict(x)

residual2 = y - yf2
```

In [6]:

```
# plot
ax2 = plt.subplot()

plt.plot(yf2,residual2,'o',color='b')

ax2.set(xlabel='Fitted Values',
        ylabel='Residual',
```
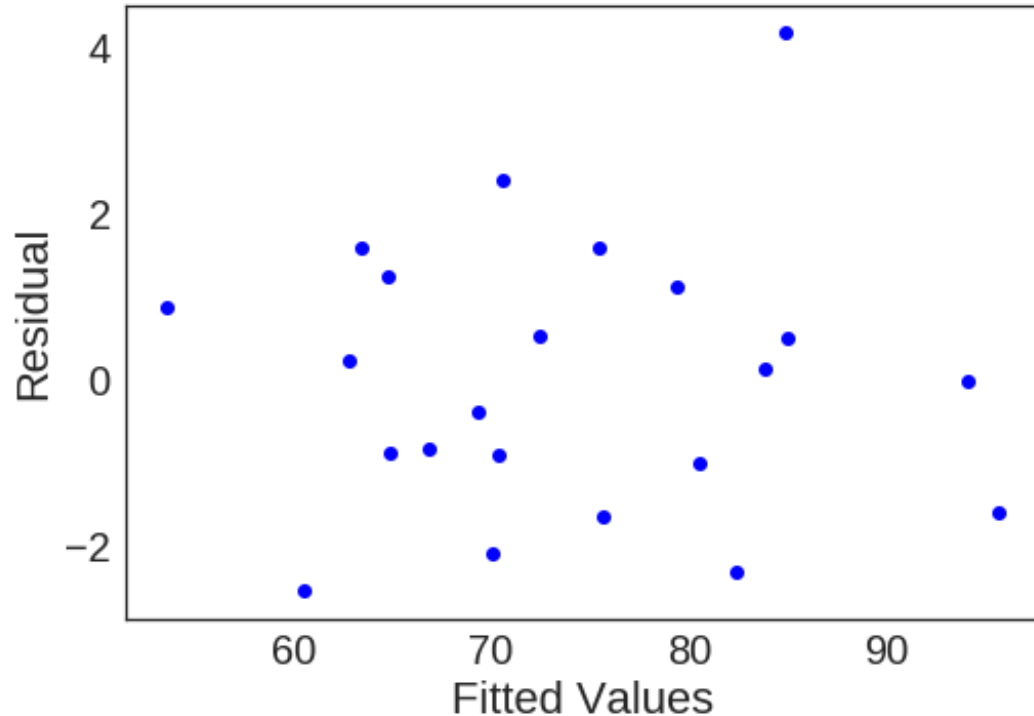
```
        title='Residual vs Fitted Values in Cube Rooted Coordinates')
```

```
[<matplotlib.text.Text at 0x7fd094660780>,
 <matplotlib.text.Text at 0x7fd09463ed68>,
 <matplotlib.text.Text at 0x7fd094607278>]
```


c) Use the plots to explain which regression is better

Both the residual plots look similar when doing regression is original and cube-root coordinates. Since both the r^2 values are above 0.9, we can find a hyperplace in high dimension that can capture the points well, regardless of the coordinate system being in original or cube-rooted.

In [ ]: