In [22]:

```python
# 导入操作系统库
import os
# 更改工作目录
os.chdir(r"D:\softwares\applied statistics\pythoncodelearning\chap1\sourcecode")
# 导入绘图库
import matplotlib.pyplot as plt
# 导入糖尿病数据集
from sklearn.datasets import load_diabetes
# 导入线性回归模型
from sklearn.linear_model import LinearRegression
# 导入回归模型评价函数，均方误差和R方
from sklearn.metrics import mean_squared_error, r2_score
# 导入绘图库中的字体管理包
from matplotlib import font_manager
# 实现中文字符正常显示
font = font_manager.FontProperties(fname=r"C:\Windows\Fonts\SimKai.ttf")
# 使用seaborn风格绘图
plt.style.use("seaborn-v0_8")

# 读取数据
diabetes_X, diabetes_y = load_diabetes(return_X_y=True, as_frame=True)
# 取出第三列
diabetes_X = diabetes_X["bmi"]
# 划分数据集，留下20个作为测试集
diabetes_X_train = diabetes_X.iloc[:-20,]
diabetes_X_test = diabetes_X.iloc[-20:,]
diabetes_y_train = diabetes_y.iloc[:-20,]
diabetes_y_test = diabetes_y.iloc[-20:,]
# 构造回归模型
regr = LinearRegression()
# 模型拟合，X必须是二维数据集
regr.fit(diabetes_X_train.values.reshape(-1,1), diabetes_y_train)
# 在测试上做预测，X必须是二维数据集
diabetes_y_pred = regr.predict(diabetes_X_test.values.reshape(-1,1))
# 估计系数
print("Coefficients: \n", regr.coef_)
# 均方误差
print("Mean squared error: {}".format(mean_squared_error(diabetes_y_test, diabet
# R2
print("Coefficient of determination: {}".format(r2_score(diabetes_y_test, diabet
# 绘制图形
fig, ax = plt.subplots(figsize=(6,6))
# 测试集的散点
ax.scatter(diabetes_X_test, diabetes_y_test, color="black")
# 预测结果线条
ax.plot(diabetes_X_test, diabetes_y_pred, color="blue", linewidth=3)
plt.show()
fig.savefig("../codeimage/code1.pdf")
```

```
Coefficients:
 [938.23786125]
Mean squared error: 2548.07239872597
Coefficient of determination: 0.47257544798227147
```