STAT 383 HW 7
Lewis Collum
Updated: April 19, 2020

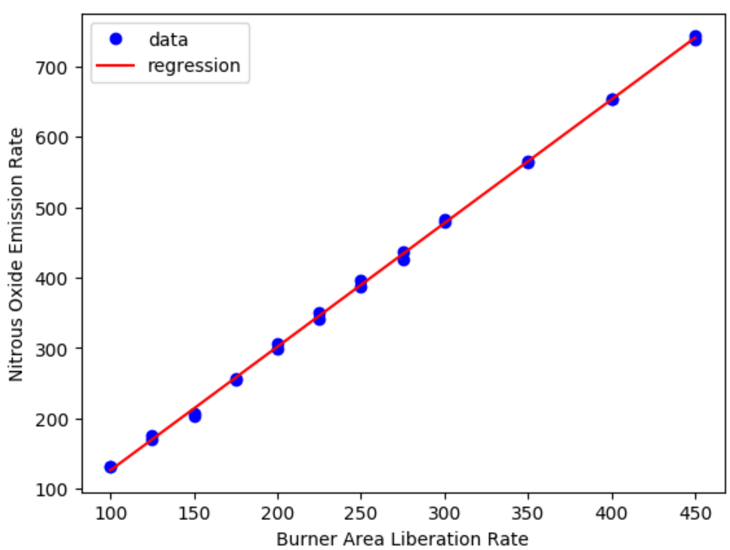
2 - Burner Rates

```
from matplotlib import pyplot
from scipy import stats
import pandas

data = pandas.read_excel('data/BurnerRates.xls')
data = data.rename(columns = {
    'Burner Area Liberation Rate': 'x',
    'Nitrous Oxide Emission Rate': 'y'})

regression = stats.linregress(data.x, data.y)
regressionLine = regression.intercept + regression.slope * data.x
```

(a)

```
print(f'Regression = '
      f'\({regression.intercept:.2f} + {regression.slope:.2f}x\)')

pyplot.plot(data.x, data.y, 'bo', label='data')
pyplot.plot(data.x, regressionLine, 'r-', label='regression')
pyplot.legend()
pyplot.xlabel('Burner Area Liberation Rate')
pyplot.ylabel('Nitrous Oxide Emission Rate')
pyplot.savefig('figure/2.png', bbox_inches='tight')
pyplot.show()
```

Regression = $-49.62 + 1.76x$
If the burner area liberation rate increases by 1, then the nitrous oxide emission rate increases by 1.76. The intercept is not directly interpretable here since the data does not consider burner area liberation rates below 100.



(b)

$rate_E$: Nitrous Oxide Emission Rate
$rate_L$: Burner Area Liberation Rate

$$rate_E = \beta_0 + \beta_1 \cdot rate_L + \epsilon$$

$$H_0 : \beta_1 = 0$$

$$H_A : \beta_1 \neq 0$$

Two-tail T test:

```
print(f'\[p = {regression.pvalue:.2E}\]')
```

$$p = 2.13E - 36$$

Since the p-value from the t-test is much less than α (0.05), we reject the null hypothesis, which indicates there is significant linear relationship between the burner area liberation rate and the nitrous oxide emission rate.

(c)

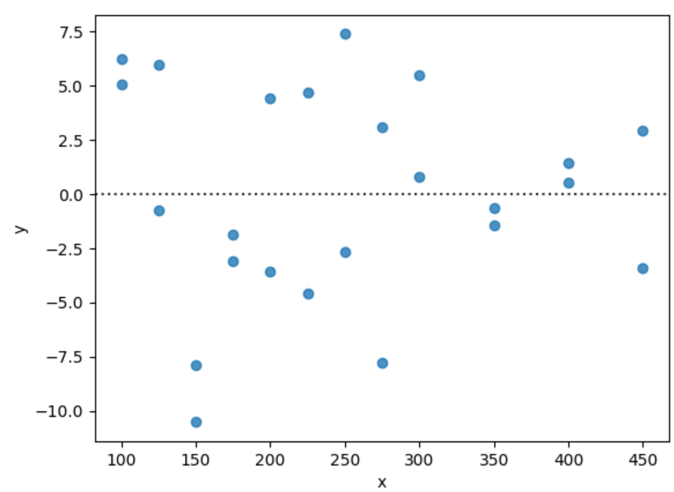
```
print(f'\[R^2 = {regression.rvalue**2:.4f}\]')
```

$$R^2 = 0.9993$$

99.97% of the variance in nitrous oxide emission rate is explained by burner area liberation rate.

(d) Residuals

```
import seaborn
seaborn.residplot(data.x, data.y)
pyplot.savefig('figure/2-resid.png')
pyplot.show()
```



There do not appear to be any outliers, and the error variance appears constant.

7 - Copper Resistivity

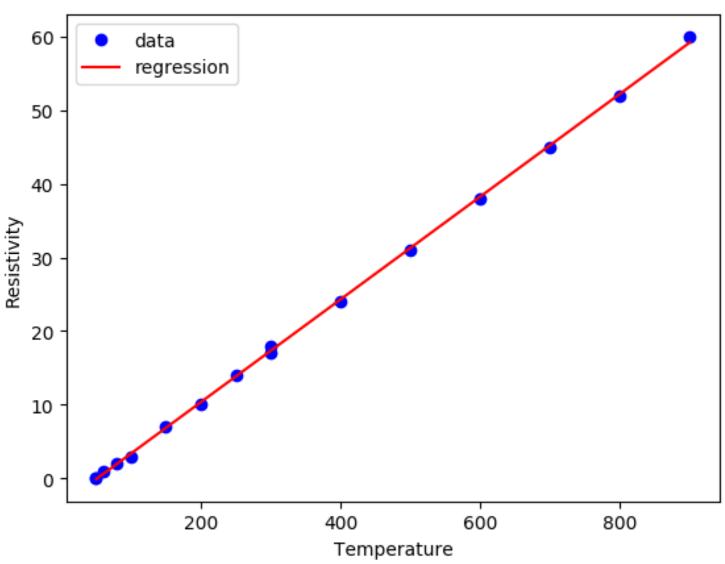
```
from matplotlib import pyplot
from scipy import stats
import pandas

data = pandas.read_excel('data/CopperResistivity.xlsx')
```

```
data = data.rename(columns = {
    'Temperature': 'x',
    'Resistivity': 'y'})

regression = stats.linregress(data.x, data.y)
regressionLine = regression.intercept + regression.slope * data.x
```

(a)

```
print(f'Regression = '
      f'\({regression.intercept:.2f} + {regression.slope:.2f}x\)')

pyplot.plot(data.x, data.y, 'bo', label='data')
pyplot.plot(data.x, regressionLine, 'r-', label='regression')
pyplot.legend()
pyplot.xlabel('Temperature')
pyplot.ylabel('Resistivity')
pyplot.savefig('figure/7.png', bbox_inches='tight')
pyplot.show()
```

Regression = $-3.61 + 0.07x$
If the temperature increases by 1 Kelvin, then the copper resistivity increases by $0.07\mathrm{n}\Omega\mathrm{m}$. The intercept is not directly interpretable here since the data does not consider temperature at 0 Kelvin.



(b)

T: Temperature
R: Resistivity

$$R = \beta_0 + \beta_1 \cdot T + \epsilon$$

$$H_0 : \beta_1 = 0$$

$$H_A : \beta_1 \neq 0$$

Two-tail T test:

```
print(f'\[p = {regression.pvalue:.2E}\]')
```

$$p = 1.50E - 25$$

Since the p-value from the t-test is much less than α (0.05), we reject the null hypothesis, which indicates there is significant linear relationship between the temperature and the resistivity of copper.

(c)

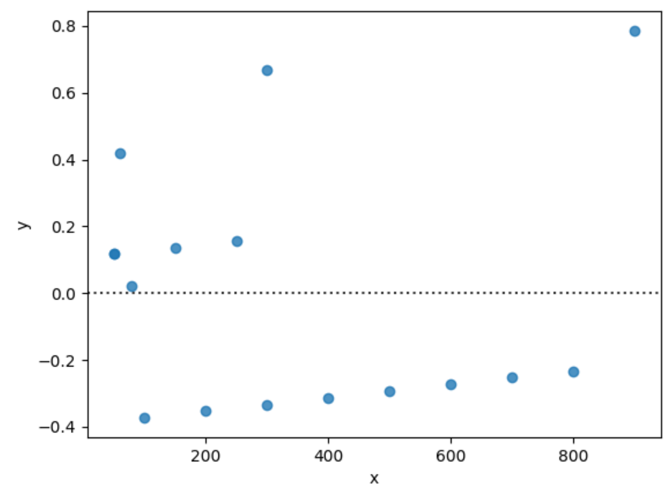
```
print(f'\[R^2 = {regression.rvalue**2:.4f}\]')
```

$$R^2 = 0.9996$$

99.96% of the variance in resistivity of copper is explained by temperature.

(d) Residuals

```
import seaborn
seaborn.residplot(data.x, data.y)
pyplot.savefig('figure/7-resid.png')
pyplot.show()
```



There do not appear to be any outliers, and the error variance appears constant. It may also be possible that there is some non-linearity at the lower ranges of temperature, and an outlier at the last temperature sample.