Coefficient Interpretation

```
# get the data
import pandas as pd

advertising = pd.read_csv('https://raw.githubusercontent.com/gitmystuff/INFO4050/main/Data
advertising.head()
```

	TV	radio	newspaper	sales
0	230.1	37.8	69.2	22.1
1	44.5	39.3	45.1	10.4
2	17.2	45.9	69.3	9.3
3	151.5	41.3	58.5	18.5
4	180.8	10.8	58.4	12.9

```
# train test split
from sklearn.model_selection import train_test_split

X_train, X_test, y_train, y_test = train_test_split(
    advertising.drop('sales', axis=1),
    advertising['sales'],
    test_size=0.25,
    random_state=42)
```

```
# create and train the model
from sklearn.linear_model import LinearRegression
from sklearn.metrics import mean_squared_error, r2_score

model = LinearRegression()
model.fit(X_train, y_train)

# test set prediction results
yhat = model.predict(X_test)
print(f'MSE: {mean_squared_error(y_true=y_test, y_pred=yhat)}')
print(f'R-Squared: {r2_score(y_test, yhat)}')
```

```
MSE: 2.880023730094192
R-Squared: 0.8935163320163658
```

```
# make a prediction
d = {'TV': 232.1, 'radio': 8.6, 'newspaper': 8.7}
d = pd.Series(d)
model.predict(pd.DataFrame([d]))
```

```
array([14.99230101])
```

```
# view the coefficients and intercept
print(list(zip(X_train, model.coef_)))
print(model.intercept_)
```

```
[('TV', 0.045433558624649886), ('radio', 0.19145653561741383), ('newspaper', 0.00256809081
2.778303460245283
```

```
# print predictions (yhat) using model.predict
yhat
```

```
array([16.38348211, 20.92434957, 21.61495426, 10.49069997, 22.17690456,
       13.02668085, 21.10309295,  7.31813008, 13.56732111, 15.12238649,
        8.92494113,  6.49924401, 14.30119928,  8.77233515,  9.58665483,
       12.09485291,  8.59621605, 16.25337881, 10.16948105, 18.85753401,
       19.5799036 , 13.15877029, 12.25103735, 21.35141984,  7.69607607,
        5.64686906, 20.79780073, 11.90951247,  9.06581044,  8.37295611,
       12.40815899,  9.89416076, 21.42707658, 12.14236853, 18.28776857,
       20.18114718, 13.99303029, 20.89987736, 10.9313953 ,  4.38721626,
        9.58213448, 12.6170249 ,  9.93851933,  8.06816257, 13.45497849,
        5.25769423,  9.15399537, 14.09552838,  8.71029827, 11.55102817])
```

The Formula

y = intercept + coef_0(TV) + coef_1(radio) + coef_2(newspaper)

```
# print predictions using formula with coefficients
print((model.intercept_ + model.coef_[0]*X_test.TV + model.coef_[1]*X_test.radio + model
```

```
[16.383482113311448, 20.924349568603073, 21.61495426261631, 10.490699965305925, 22.1769045
```

Interpreting a coefficient: $1000 dollars on radio advertising would be associated with an increase of sales by 0.19 * 1000, or 190 units, given spending stays the same.

```
# add constant and build model
import statsmodels.api as sm
import statsmodels.formula.api as smf

X_train.insert(0, 'const', 1)
model = sm.OLS(y_train, X_train).fit()
model.summary()
```

OLS Regression Results

Dep. Variable:	sales	**R-squared:**	0.897
Model:	OLS	**Adj. R-squared:**	0.895
Method:	Least Squares	**F-statistic:**	422.2
Date:	Fri, 30 Jun 2023	**Prob (F-statistic):**	1.02e-71
Time:	13:00:28	**Log-Likelihood:**	-289.20
No. Observations:	150	**AIC:**	586.4
Df Residuals:	146	**BIC:**	598.4
Df Model:	3		
Covariance Type:	nonrobust		

	coef	**std err**	**t**	**P>|t|**	**[0.025**	**0.975]**
const	2.7783	0.375	7.415	0.000	2.038	3.519
TV	0.0454	0.002	27.960	0.000	0.042	0.049
radio	0.1915	0.010	19.076	0.000	0.172	0.211
newspaper	0.0026	0.007	0.356	0.722	-0.012	0.017

Omnibus:	61.891	**Durbin-Watson:**	2.153
Prob(Omnibus):	0.000	**Jarque-Bera (JB):**	206.420
Skew:	-1.581	**Prob(JB):**	1.50e-45
Kurtosis:	7.799	**Cond. No.**	485.

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.

```
print(0.0454 + .002 * 2)
print(0.0454 / .002)
print(1.0250   /   0.033)
# 31.191
```

```
0.0494
22.7
31.060606060606055
```

```
# view the coefficients and intercept
print(model.params)
```

```
const        2.778303
TV           0.045434
radio        0.191457
newspaper    0.002568
dtype: float64
```

If we want to see an increase of 20 units in sales, given a unit is one million dollars, and we are spending $37,800 on radio advertising, how much would we need to spend on TV advertising? Note: our sales has been scaled to match units of TV and radio spending, given TV and radio units are $1000 each

Consider this equation:

$$y = \beta_0 + \beta_1(X_1) + \beta_2(X_2)$$

To see an increase of 20 units in sales knowing that we are spending $37,800 on radio advertising, what do we need to spend on TV?

```
# solve for X1
intercept = model.params['const']
B1 = model.params['TV']
B2 = model.params['radio']
X2 = 37.8
print(f'y = {intercept:0.2f} +( {B1:0.2f} * X1) + ({B2:0.2f} * {X2})')
```

```
y = 2.78 +( 0.05 * X1) + (0.19 * 37.8)
```