Model

$$\hat{y} = 11.9524 + 0.0029 \cdot \hat{x}, \quad \text{where} \quad \hat{y} = \ln(\text{price}) \text{ and } \hat{x} = \text{sqft\_living}^{0.78}$$

Explanation of the Model

Intercept and slope

The model has the sample intercept of 11.9524 and the slope of 0.0029. To interpret the slope, we have to transform $\hat{x}$ and $\hat{y}$ towards original *sqft_living* and *price*. Let $x = sqft\_living$ and $y = price$ for the derivation. We have $\hat{x} = x^{0.78}$ and $\hat{y} = \ln(y) \implies y = e^{\hat{y}}$. Suppose the original sqft_living is x_1 and it moved up to x_2, then we have the following :

$$\begin{aligned} y_1 =& e^{\hat{y}_1} = e^{11.9524+0.0029 \cdot x_1^{0.78}} = e^{11.9524} \cdot e^{0.0029 \cdot x_1^{0.78}} \\ y_2 =& e^{\hat{y}_2} = e^{11.9524+0.0029 \cdot x_2^{0.78}} = e^{11.9524} \cdot e^{0.0029 \cdot x_2^{0.78}} \end{aligned}$$

To understand the change in *price* in percents we will use the following formula:

$$\begin{aligned} 100 \times \left[\frac{y_2 - y_1}{y_1}\right] =& 100 \times \left[\frac{y_2}{y_1} - 1\right] = 100 \times \left[\frac{\cancel{e^{11.9524}} \cdot e^{0.0029 \cdot x_2^{0.78}}}{\cancel{e^{11.9524}} \cdot e^{0.0029 \cdot x_1^{0.78}}} - 1\right] \\ =& 100 \times \left[e^{0.0029\left(x_2^{0.78} - x_1^{0.78}\right)} - 1\right] \end{aligned}$$

For example, if the *sqft_living* is $1000ft$ and we increase it to $1100ft$, we will get the change in price of

$$100 \times \left[e^{0.0029(1100^{0.78} - 1000^{0.78})} - 1\right] \approx 5.02\%.$$

In this particular example 10% change in sqft_living starting from $x_1 = 1000ft$ forces 5.02% change in price.

R^2

The model has $R^2 \approx 0.45$. This means that our model explains about 45% of the variation by using *sqft_living* as independent variable.

ANOVA

Is our model with one explanatory variable better than the model with zero explanatory variables?
Our model has $F - statistic = 1.737 \times 10^4$ and $Prob > F$ is 0.000.
The Null Hypothesis: The slope$= 0$
The Alternative Hypothesis: The slope$\neq 0$
Our p-value for this model is $p = 0.000 < 0.05 = \alpha$. Thus, we have enough evidence to reject the Null Hypothesis at 5% level of significance and we conclude that the Test tells us, that our slope is not 0. Since our p-value is 0, there is a 0% probability that the improvements that we are seeing with our one independent variable model are due to random chance alone.