

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

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 $x = \text{sqft_living}$ and $y = \text{price}$ 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:

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

For example, if the *sqft_living* is 1000*ft* and we increase it to 1100, we will get 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 $x_1 = 1000\text{ft}$ forces 5.02% change in price.

Check Statistical Hypotheses of the Regression

Linearity:

The Null Hypothesis: The model is linearly predicted by the feature,

The Alternative Hypothesis: The model is not linearly predicted by the feature.

Our p-value for this model is $p = 0.877 > 0.05 = \alpha$. Thus, we don't have enough evidence to reject **The Null Hypothesis** and we conclude that our model satisfies Linearity Assumption.

Normality Assumption for Errors

To check Normality, I used the following checks:

1. QQ-Plot

