

# Class Example

Alan T. Arnholt

Last compiled: March 07, 2023 at 01:10:53 PM

## 1 Code chunks (hidden)

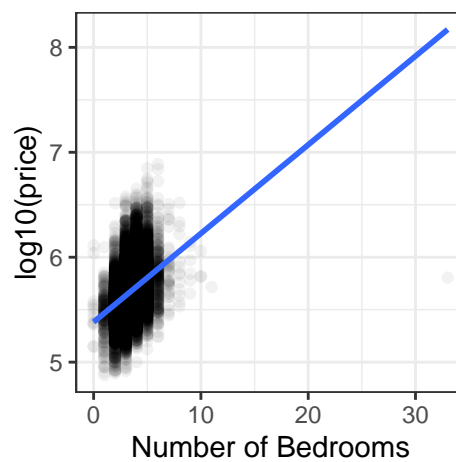


Figure 1: Scatterplot of Log10 of Price versus Number of Bedrooms

### 1.1 Data munging

```
# removing outlier
house_prices %>%
  filter(bedrooms < 33) -> house_prices
```

### 1.2 Scatterplot without and with outlier removed

```
p2 <- ggplot(data = house_prices, aes(x = bedrooms, y = log10(price))) +
  geom_point(alpha = 0.05) +
  geom_smooth(method = "lm", se = FALSE) +
  theme_bw() +
  labs(x = "Number of Bedrooms")
library(patchwork)
p1 + p2
```

The left graph in Figure 2 contains all of the data in `house_prices` while the right graph in Figure 2 removes the outlier in Figure 1 and creates a scatterplot without the “outlier”.

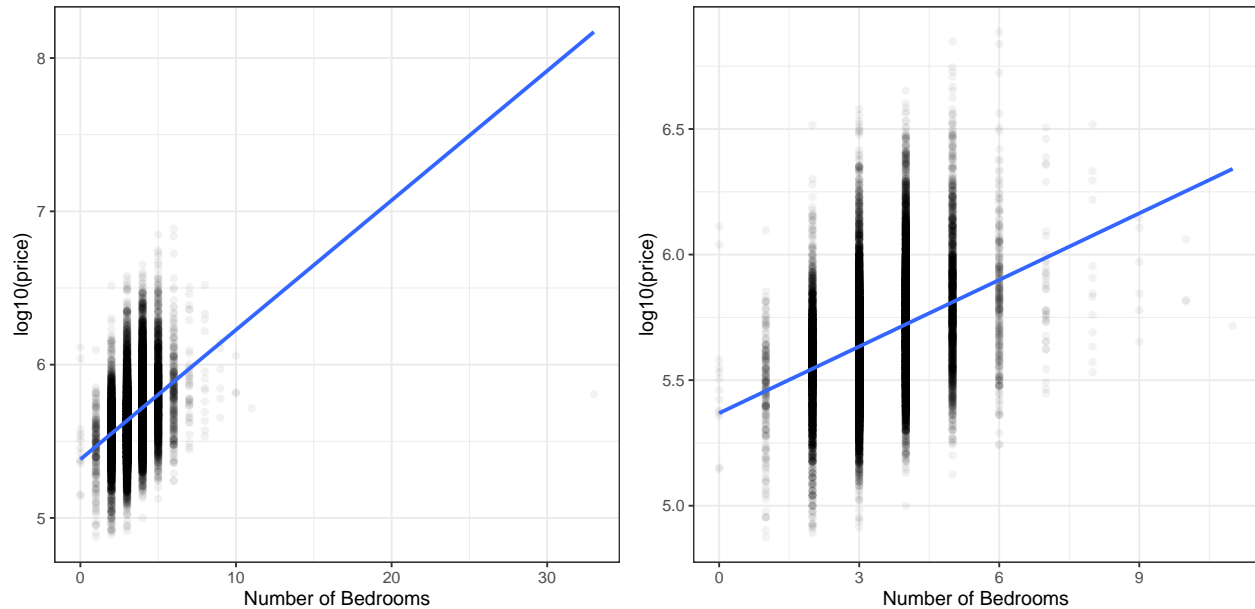


Figure 2: Scatterplots with and without outlier for Log10 Price versus Number of Bedrooms

Table 1: A different formatting of T1

| term               | estimate | std_error | statistic | p_value | lower_ci | upper_ci |
|--------------------|----------|-----------|-----------|---------|----------|----------|
| intercept          | 2.693    | 0.023     | 116.373   | 0       | 2.647    | 2.738    |
| log10(sqft_living) | 0.941    | 0.008     | 117.757   | 0       | 0.925    | 0.957    |
| bedrooms           | -0.033   | 0.002     | -20.505   | 0       | -0.036   | -0.030   |

## 2 Creating models and showing regression output

```
mod1 <- lm(log10(price) ~ log10(sqft_living) + bedrooms, data = house_prices)
get_regression_table(mod1) -> T1
T1
```

```
# A tibble: 3 x 7
  term          estimate std_error statistic p_value lower_ci upper_ci
<chr>         <dbl>    <dbl>    <dbl>   <dbl>   <dbl>   <dbl>
1 intercept      2.69      0.023     116.      0      2.65     2.74
2 log10(sqft_living) 0.941    0.008     118.      0      0.925    0.957
3 bedrooms     -0.033    0.002    -20.5      0     -0.036   -0.03
```

Note: the output from T1 could be made to look better. Consider the `kable()` function from `knitr`:

```
knitr::kable(T1, caption = "A different formatting of T1")
```

### 2.1 Writing the regression equation with $\LaTeX$ and inline R code

The least squares regression equation from regressing  $\log_{10}$  of price onto  $\log_{10}$  of `sqft_living` and `bedrooms` is written in Equation (1).

$$\widehat{\log_{10}(\text{price})} = 2.693 + 0.941 \cdot \log_{10}(\text{sqft\_living}) - 0.033 \cdot \text{bedrooms} \quad (1)$$