

# Lab 3

Yunting Chiu

2021-02-05

## R Lab 2. Review of T-tests and F-tests

```
library(tidyverse)
```

```
## -- Attaching packages --- tidyverse 1.3.0 --
```

```
## v ggplot2 3.3.2      v purrr  0.3.4
```

```
## v tibble  3.0.3      v dplyr  1.0.2
```

```
## v tidyr   1.1.2      v stringr 1.4.0
```

```
## v readr   1.3.1      v forcats 0.5.0
```

```
## -- Conflicts ----- tidyverse_conflicts() --
```

```
## x dplyr::filter() masks stats::filter()
```

```
## x dplyr::lag()    masks stats::lag()
```

```
H <- read_csv("HOME_SALES.csv")
```

```
## Parsed with column specification:
```

```
## cols(
```

```
##   ID = col_double(),
```

```
##   SALES_PRICE = col_double(),
```

```
##   FINISHED_AREA = col_double(),
```

```
##   BEDROOMS = col_double(),
```

```
##   BATHROOMS = col_double(),
```

```
##   GARAGE_SIZE = col_double(),
```

```
##   YEAR_BUILT = col_double(),
```

```
##   STYLE = col_double(),
```

```
##   LOT_SIZE = col_double(),
```

```
##   AIR_CONDITIONER = col_character(),
```

```
##   POOL = col_character(),
```

```
##   QUALITY = col_character(),
```

```
##   HIGHWAY = col_character()
```

```
## )
```

```
head(H)
```

```
## # A tibble: 6 x 13
```

```
##   ID SALES_PRICE FINISHED_AREA BEDROOMS BATHROOMS GARAGE_SIZE YEAR_BUILT
```

```
##   <dbl>      <dbl>      <dbl>    <dbl>    <dbl>      <dbl>    <dbl>
```

```
## 1     1        360        3032      4        4          2      1972
```

```
## 2     2        340        2058      4        2          2      1976
```

```
## 3     3        250        1780      4        3          2      1980
```

```
## 4     4        206.        1638      4        2          2      1963
```

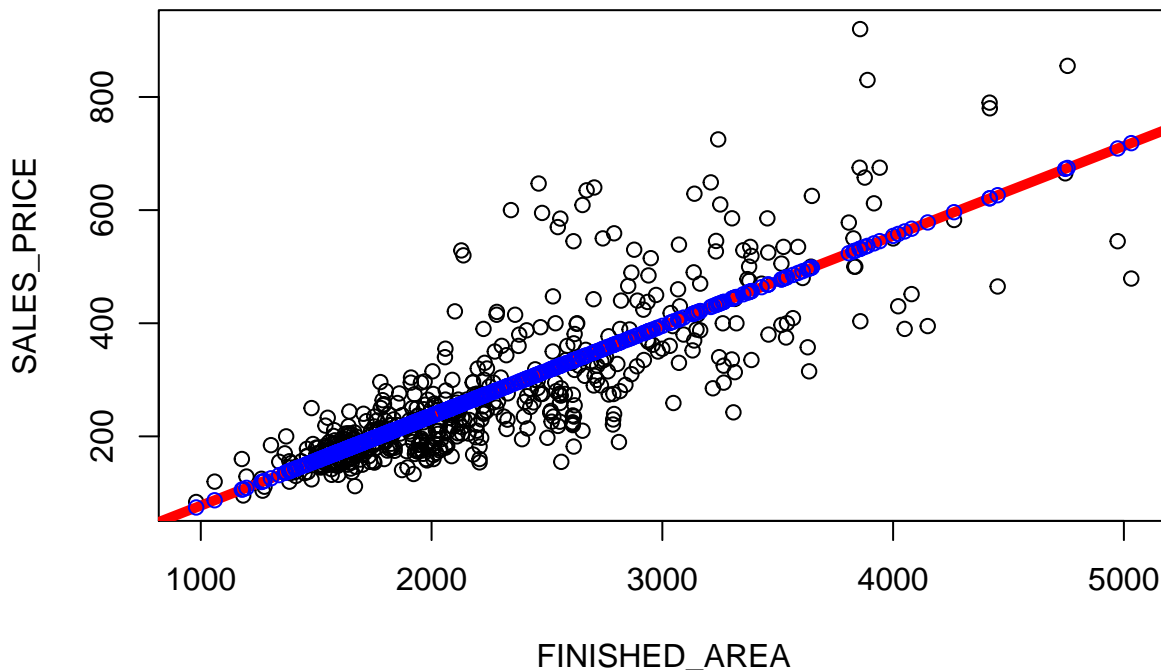
```
## 5     5        276.        2196      4        3          2      1968
```

```
## 6      6      248      1966      4      3      5      1972
## # ... with 6 more variables: STYLE <dbl>, LOT_SIZE <dbl>,
## #   AIR_CONDITIONER <chr>, POOL <chr>, QUALITY <chr>, HIGHWAY <chr>
```

```
attach(H)
names(H)
```

```
## [1] "ID"          "SALES_PRICE"  "FINISHED_AREA" "BEDROOMS"
## [5] "BATHROOMS"   "GARAGE_SIZE"  "YEAR_BUILT"    "STYLE"
## [9] "LOT_SIZE"    "AIR_CONDITIONER" "POOL"          "QUALITY"
## [13] "HIGHWAY"
```

```
plot(FINISHED_AREA, SALES_PRICE)
reg = lm(SALES_PRICE ~ FINISHED_AREA) #predict SALES_PRICE based on FINISHED_AREA
abline(reg,col="red",lwd=5)
Yhat = predict(reg, x=FINISHED_AREA) #red line = Yhat = prediction
points(FINISHED_AREA, Yhat, col="blue")
```



```
predict(reg,data.frame(FINISHED_AREA=c(2500,4000,6000)))
```

```
##      1      2      3
## 315.9426 554.3680 872.2684
```

```
summary(reg)
```

```
##
## Call:
## lm(formula = SALES_PRICE ~ FINISHED_AREA)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -239.40  -39.84   -7.64   23.52  388.36
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
##
```

```
## (Intercept)  -81.432946  11.551846  -7.049 5.74e-12 ***
## FINISHED_AREA  0.158950   0.004875  32.605  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 79.12 on 520 degrees of freedom
## Multiple R-squared:  0.6715, Adjusted R-squared:  0.6709
## F-statistic: 1063 on 1 and 520 DF,  p-value: < 2.2e-16
```

```
anova(reg)
```

```
## Analysis of Variance Table
##
## Response: SALES_PRICE
##              Df Sum Sq Mean Sq F value    Pr(>F)
## FINISHED_AREA   1 6655486 6655486  1063.1 < 2.2e-16 ***
## Residuals     520 3255426    6260
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```