# Homework Module 3

## Tyler Black

## Due Date: 17-07-2024

### Loading dataset and doing Summary Statistics

library(tidyverse)

## ── Attaching core tidyverse packages ──────────────────────── tidyverse 2.0.0 ──  
## ✔ dplyr 1.1.4 ✔ readr 2.1.5  
## ✔ forcats 1.0.0 ✔ stringr 1.5.1  
## ✔ ggplot2 3.5.1 ✔ tibble 3.2.1  
## ✔ lubridate 1.9.3 ✔ tidyr 1.3.1  
## ✔ purrr 1.0.2   
## ── Conflicts ────────────────────────────────────────── tidyverse\_conflicts() ──  
## ✖ dplyr::filter() masks stats::filter()  
## ✖ dplyr::lag() masks stats::lag()  
## ℹ Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors

library(readr)  
diamonds\_updated <- read\_csv("diamonds\_updated.csv")

## Rows: 308 Columns: 5  
## ── Column specification ────────────────────────────────────────────────────────  
## Delimiter: ","  
## dbl (5): IDNO, WEIGHT, COLOR, CLARITY, PRICE  
##   
## ℹ Use `spec()` to retrieve the full column specification for this data.  
## ℹ Specify the column types or set `show\_col\_types = FALSE` to quiet this message.

summary(diamonds\_updated)

## IDNO WEIGHT COLOR CLARITY   
## Min. : 1.00 Min. :0.1800 Min. :1.00 Min. :1.000   
## 1st Qu.: 77.75 1st Qu.:0.3500 1st Qu.:3.00 1st Qu.:2.000   
## Median :154.50 Median :0.6200 Median :4.00 Median :3.000   
## Mean :154.50 Mean :0.6309 Mean :3.75 Mean :3.153   
## 3rd Qu.:231.25 3rd Qu.:0.8500 3rd Qu.:5.00 3rd Qu.:4.000   
## Max. :308.00 Max. :1.1000 Max. :6.00 Max. :5.000   
## PRICE   
## Min. : 638   
## 1st Qu.: 1625   
## Median : 4215   
## Mean : 5019   
## 3rd Qu.: 7446   
## Max. :16008

### Correlation Analysis

cor(diamonds\_updated,method="pearson")

## IDNO WEIGHT COLOR CLARITY PRICE  
## IDNO 1.0000000 0.3432758 0.11790297 -0.25025045 0.37111454  
## WEIGHT 0.3432758 1.0000000 0.11773671 0.33733803 0.94472662  
## COLOR 0.1179030 0.1177367 1.00000000 -0.09560535 -0.07822459  
## CLARITY -0.2502504 0.3373380 -0.09560535 1.00000000 0.19829033  
## PRICE 0.3711145 0.9447266 -0.07822459 0.19829033 1.00000000

### Linear Regression

model <- lm(PRICE ~ CLARITY + COLOR + WEIGHT, data = diamonds\_updated)  
summary(model)

##   
## Call:  
## lm(formula = PRICE ~ CLARITY + COLOR + WEIGHT, data = diamonds\_updated)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -1072.0 -494.5 -203.9 380.2 4260.4   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 422.71 171.57 2.464 0.0143 \*   
## CLARITY -443.85 34.89 -12.723 <2e-16 \*\*\*  
## COLOR -522.40 30.51 -17.123 <2e-16 \*\*\*  
## WEIGHT 12608.90 163.22 77.253 <2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 736.6 on 304 degrees of freedom  
## Multiple R-squared: 0.9536, Adjusted R-squared: 0.9532   
## F-statistic: 2083 on 3 and 304 DF, p-value: < 2.2e-16

vcov(model)

## (Intercept) CLARITY COLOR WEIGHT  
## (Intercept) 29434.578 -3147.7450 -3473.4027 -7487.8310  
## CLARITY -3147.745 1217.0409 154.0698 -2007.9862  
## COLOR -3473.403 154.0698 930.8105 -797.0349  
## WEIGHT -7487.831 -2007.9862 -797.0349 26639.4702