Spring 2025 ..... Score _____ / 20

Homework #2

[Chaemin Lee, lab2 Anthony]

---

highlight and click "Run" the line below before knitting

install.packages("rmarkdown")

```
# set seed replace 12345678 with your student ID
seed = 12345678
```

```
# loads in data for the full population
pop<-read.csv("HW2.csv")
names(pop) <- c("G", "X", "Y")
```

```
# sets the seed for the random number generator
set.seed(seed+25)
```

```
# assigns a "random" sample of 10 from group A and 5 from group B from the population
G_A<-pop[sample(nrow(pop), 10, pop$G == "A", replace=FALSE),]
G_B<-pop[sample(nrow(pop), 5, pop$G == "B", replace=FALSE),]
```

```
# use this data
data <- rbind(G_A, G_B)
data
```

```
##     G  X  Y
## 329 A  7 10
## 501 A 17 15
## 337 A 10  7
## 268 A 20 17
## 622 A  5  7
## 746 A 17 15
## 113 A  7 10
## 192 A 20 17
## 870 A 16 16
## 150 A  2  2
## 160 B  8 18
## 291 B  8 11
## 971 B  8 11
## 702 B  8 11
## 348 B  6 10
```

```
# regression for A and B combined
model_full <- lm(Y ~ X, data=data)
summary(model_full)
```

```
##
## Call:
## lm(formula = Y ~ X, data = data)
##
```

```
## Residuals:
##     Min      1Q  Median      3Q     Max
## -4.5252 -0.7254  0.4080  0.7947  7.7947
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   5.2985     1.6005   3.310 0.005633 **
## X             0.6133     0.1336   4.591 0.000506 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.889 on 13 degrees of freedom
## Multiple R-squared:  0.6185, Adjusted R-squared:  0.5892
## F-statistic: 21.08 on 1 and 13 DF,  p-value: 0.0005059
```

```
# regression for A
model_A <- lm(Y ~ X, data=G_A)
summary(model_A)
```

```
##
## Call:
## lm(formula = Y ~ X, data = G_A)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.0582 -0.4001 -0.1975  1.3057  2.1444
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  2.71630    1.24525   2.181   0.0607 .
## X            0.73419    0.09128   8.043  4.2e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.819 on 8 degrees of freedom
## Multiple R-squared:  0.8899, Adjusted R-squared:  0.8762
## F-statistic: 64.69 on 1 and 8 DF,  p-value: 4.201e-05
```

```
# regression for B
model_B <- lm(Y ~ X, data=G_B)
summary(model_B)
```

```
##
## Call:
## lm(formula = Y ~ X, data = G_B)
##
## Residuals:
##        160        291        971        702        348
##  5.250e+00 -1.750e+00 -1.750e+00 -1.750e+00  6.661e-15
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.750     14.952   0.117    0.914
## X              1.375      1.957   0.703    0.533
##
## Residual standard error: 3.5 on 3 degrees of freedom
```

```
## Multiple R-squared:  0.1414, Adjusted R-squared:  -0.1449
## F-statistic: 0.4939 on 1 and 3 DF,  p-value: 0.5328
```

```r
# creates plot
plot(data$X, data$Y, main=c(paste("Scatterplot by Subgroups")), xlab="X", ylab="Y", xlim=c(0,30), ylim=c
abline(model_full)
abline(model_A, lty = "dashed")
abline(model_B, lty = "dotted")
legend("bottomright", pch=c(1,3), c("A","B"), bty="o", cex=.8)
```