```
> # Likhit Garimella
> # Regression Analysis HW-3
>
> # libraries
> #install.packages("readr")
> library(readr)
>
> # importing dataset
> data <- read_csv('/Users/likhitgarimella/Desktop/SummerSemester/
B1.csv',show_col_types = FALSE)
> data
# A tibble: 28 × 10
       y    x1    x2    x3    x4    x5    x6    x7    x8    x9
   <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>
 1    10  2113  1985  38.9  64.7     4   868  59.7  2205  1917
 2    11  2003  2855  38.8  61.3     3   615  55    2096  1575
 3    11  2957  1737  40.1  60      14   914  65.6  1847  2175
 4    13  2285  2905  41.6  45.3    -4   957  61.4  1903  2476
 5    10  2971  1666  39.2  53.8    15   836  66.1  1457  1866
 6    11  2309  2927  39.7  74.1     8   786  61    1848  2339
 7    10  2528  2341  38.1  65.4    12   754  66.1  1564  2092
 8    11  2147  2737  37    78.3    -1   761  58    1821  1909
 9     4  1689  1414  42.1  47.6    -3   714  57    2577  2001
10     2  2566  1838  42.3  54.2    -1   797  58.9  2476  2254
# ℹ 18 more rows
# ℹ Use `print(n = ...)` to see more rows
>
> # creating data frame with the columns of interest
> df <- data.frame(y = data$y, x2 = data$x2, x7 = data$x7, x8 =
data$x8)
>
> # fit the multiple regression model
> model <- lm(y ~ x2 + x7 + x8, data = data)
> model_summary <- summary(model)
>
> # model summary
> summary(model)

Call:
lm(formula = y ~ x2 + x7 + x8, data = data)

Residuals:
    Min      1Q  Median      3Q     Max
-3.0370 -0.7129 -0.2043  1.1101  3.7049

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) -1.808372   7.900859  -0.229 0.820899
x2           0.003598   0.000695   5.177 2.66e-05 ***
x7           0.193960   0.088233   2.198 0.037815 *
x8          -0.004816   0.001277  -3.771 0.000938 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

```
Residual standard error: 1.706 on 24 degrees of freedom
Multiple R-squared:  0.7863,    Adjusted R-squared:  0.7596
F-statistic: 29.44 on 3 and 24 DF,  p-value: 3.273e-08

>
> # (a) calculating confidence intervals
> confidence_interval <- confint(model, "x7", level = 0.95)
> confidence_interval
        2.5 %    97.5 %
x7 0.01185532 0.3760651
>
> # (b) defining values of x2, x7, and x8
> x2 <- 2300
> x7 <- 56.0
> x8 <- 2100
>
> # new data frame with the specific values
> new_data <- data.frame(x2 = x2, x7 = x7, x8 = x8)
>
> # using the model to predict the confidence interval of number of
games won
> pred_conf <- predict(model, newdata = new_data, interval =
"confidence", level = 0.95)
> pred_conf
       fit      lwr      upr
1 7.216424 6.436203 7.996645
>
> # (c) using the model to predict the prediction interval number of
games won
> predictions_pred <- predict(model, newdata = new_data, interval =
"prediction", level = 0.95)
> predictions_pred
       fit      lwr      upr
1 7.216424 3.609523 10.82332
>
> # 3.3 (d) interpretation explanation
> #
> # a) The 95% confidence interval for β7 is (0.01185532,
0.3760651).
> # This interval suggests that there is strong evidence to support
the hypothesis that the variable x7 has a positive and significant
impact on the number of games won by a team.
> # We can be 95% confident that the true population value of the
coefficient β7 lies within this interval.
> #
> # b) The 95% confidence interval for the mean number of games won,
when x2 = 2300, x7 = 56.0, and x8 = 2100, is (6.436203, 7.996645).
> # This interval indicates that if we were to conduct the
experiment multiple times with teams having these specific values of
x2, x7, and x8, 95% of the resulting confidence intervals would
contain the true population mean number of games won.
> # In practical terms, it provides a range within which we can be
95% confident that the actual mean number of games won falls.
> #
```

```
> # c) The 95% prediction interval for the number of games won is
(3.609523, 10.82332).
> # This prediction interval accounts for both the uncertainty
associated with the regression model and the variability of
individual observations.
> # For a specific team with x2 = 2300, x7 = 56.0, and x8 = 2100, we
can be 95% confident that the actual number of games won will fall
within this interval.
> # In other words, the prediction interval provides a range within
which we expect the true value to lie with a certain level of
confidence.
> #
>
```