Lab 3 - MLR: Part 1

PrabhTalwar

2022-10-16

```
library(readr)
```

```
## Warning: package 'readr' was built under R version 4.1.3
```

```
library(tidyverse)
```

```
## Warning: package 'tidyverse' was built under R version 4.1.3
```

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.1 --
```

```
## v ggplot2 3.3.6     v dplyr   1.0.9
## v tibble  3.1.7     v stringr 1.4.0
## v tidyr   1.2.0     v forcats 0.5.1
## v purrr   0.3.4
```

```
## Warning: package 'ggplot2' was built under R version 4.1.3
```

```
## Warning: package 'tibble' was built under R version 4.1.3
```

```
## Warning: package 'tidyr' was built under R version 4.1.3
```

```
## Warning: package 'purrr' was built under R version 4.1.3
```

```
## Warning: package 'dplyr' was built under R version 4.1.3
```

```
## Warning: package 'stringr' was built under R version 4.1.3
```

```
## Warning: package 'forcats' was built under R version 4.1.3
```

```
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
teamperf <- read_csv("teamperf.csv")
```

```
## Rows: 23 Columns: 5
```

```
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## dbl (5): team, intrapersonal_range, stress_range, mood_range, project_score
##
## 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.
```

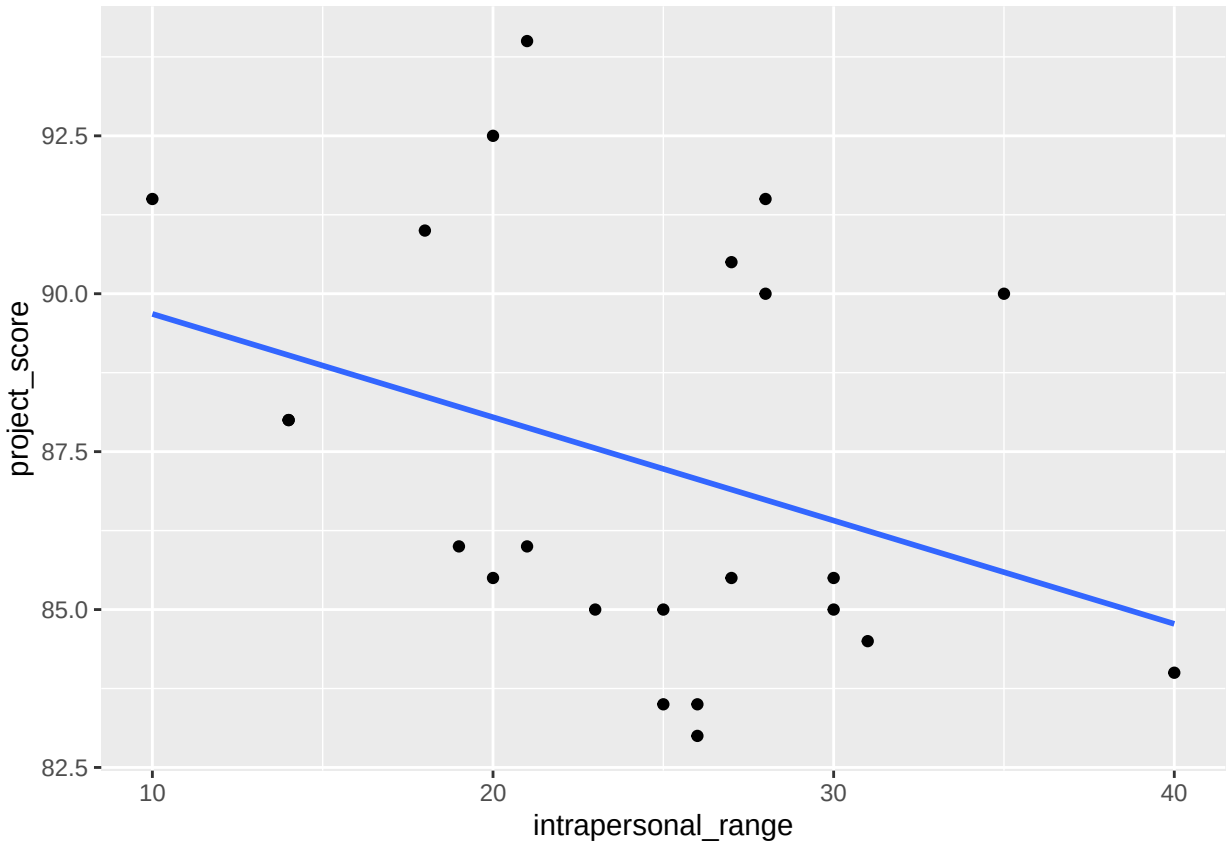
(a) Hypothesize a first-order model for project scores y as a function of x1, x2 and x3

Our dependent variable(y) is mean project score. Three independent variables for each team: range of intra-personal scores (x1), range of stress management (x2),and range of mood scores (x3)

(b) Use a appropriate plots to verify a linear relationship

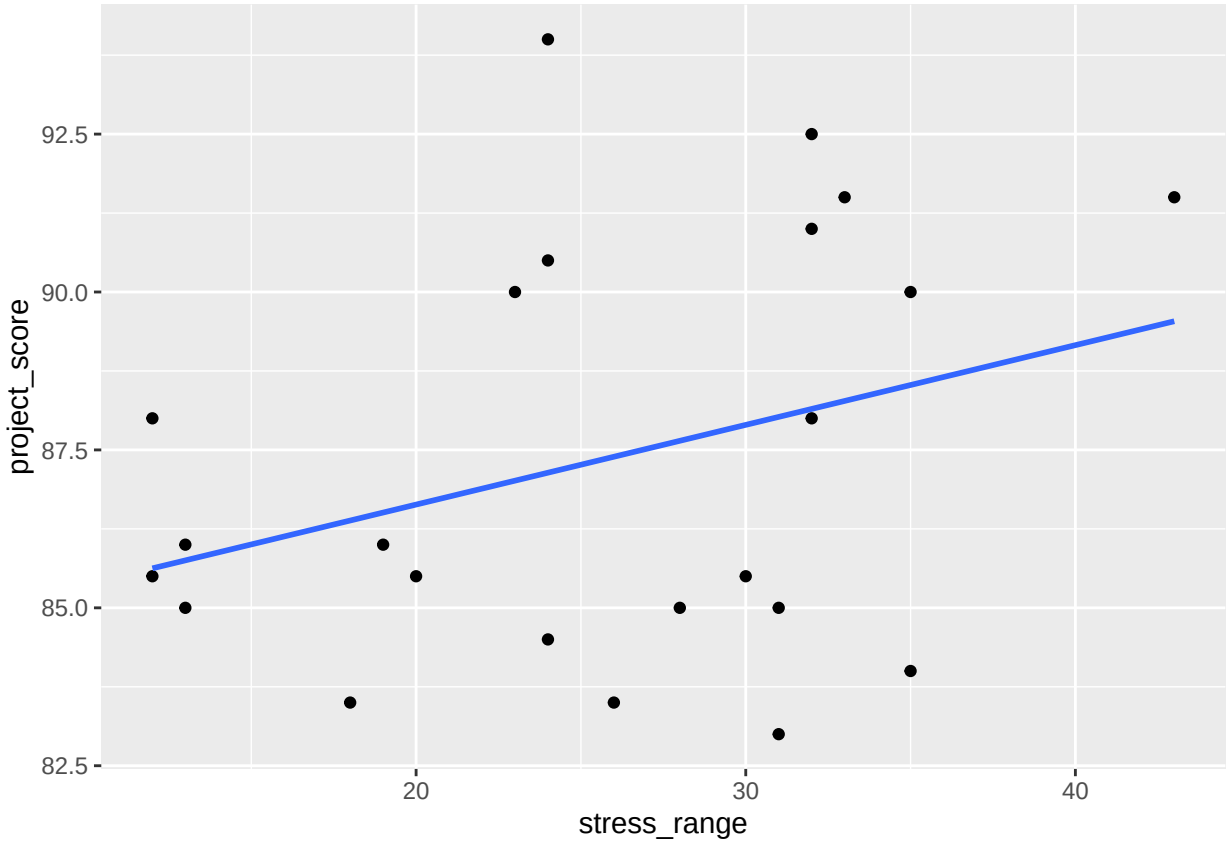
```
ggplot(data = teamperf, mapping = aes(x = intrapersonal_range, y = project_score))+
  geom_point()+
  geom_smooth(method='lm', se = FALSE)
```

```
## `geom_smooth()` using formula 'y ~ x'
```



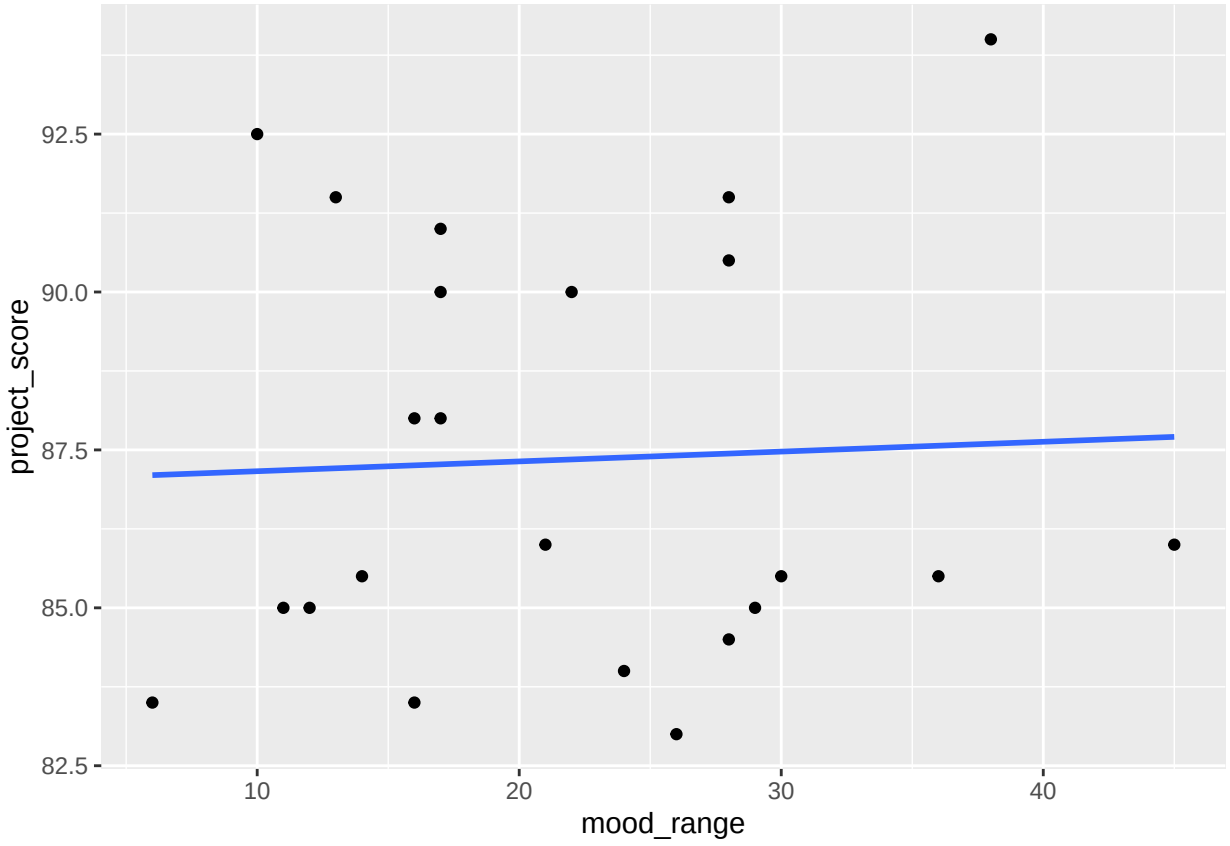

```
ggplot(data = teamperf, mapping = aes(x = stress_range, y = project_score))+
  geom_point()+
  geom_smooth(method='lm', se = FALSE)
```

```
## `geom_smooth()` using formula 'y ~ x'
```




```
ggplot(data = teamperf, mapping = aes(x = mood_range, y = project_score))+
  geom_point()+
  geom_smooth(method='lm', se = FALSE)
```

```
## `geom_smooth()` using formula 'y ~ x'
```



(c) Fit the model in part (a), to the data using R

```
model = lm(project_score ~ intrapersonal_range + stress_range + mood_range, data = teamperf)
coef(model)
```

```
##         (Intercept) intrapersonal_range        stress_range          mood_range
##         86.64143382         -0.19937014          0.15605653          0.07028216
```

(d) Use matrices to verify the equations and produce the same coefficients.

```
n = nrow(teamperf)
p = length(coef(model))

X = cbind(rep(1, n), teamperf$intrapersonal_range, teamperf$stress_range, teamperf$mood_range)
y = teamperf$project_score
beta_hat = solve(t(X) %*% X) %*% t(X) %*% y
beta_hat
```

```
##              [,1]
## [1,] 86.64143382
## [2,] -0.19937014
```

```
## [3,]  0.15605653
## [4,]  0.07028216
```

(e) Is there sufficient evidence to indicate the overall model is statistically useful for predicting y? Test using alpha = 0.10.

```
summary(model)
```

```
##
## Call:
## lm(formula = project_score ~ intrapersonal_range + stress_range +
##     mood_range, data = teamperf)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -5.1229 -1.8920 -0.8116  1.7606  5.1293
##
## Coefficients:
##                     Estimate Std. Error t value Pr(>|t|)
## (Intercept)         86.64143    3.26033  26.574   <2e-16 ***
## intrapersonal_range -0.19937    0.09537  -2.090   0.0503 .
## stress_range         0.15606    0.07835   1.992   0.0610 .
## mood_range           0.07028    0.06845   1.027   0.3174
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.023 on 19 degrees of freedom
## Multiple R-squared:  0.2816, Adjusted R-squared:  0.1681
## F-statistic: 2.482 on 3 and 19 DF,  p-value: 0.09204
```

The p-value for our model is 0.09204 which is less than 0.10, means that our model is statistically significant.

(f) Evaluate the model using statistics R2 adjusted and 2s.

```
summary(model)
```

```
##
## Call:
## lm(formula = project_score ~ intrapersonal_range + stress_range +
##     mood_range, data = teamperf)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -5.1229 -1.8920 -0.8116  1.7606  5.1293
##
## Coefficients:
##                     Estimate Std. Error t value Pr(>|t|)
## (Intercept)         86.64143    3.26033  26.574   <2e-16 ***
## intrapersonal_range -0.19937    0.09537  -2.090   0.0503 .
```

```
## stress_range         0.15606    0.07835   1.992   0.0610 .
## mood_range           0.07028    0.06845   1.027   0.3174
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.023 on 19 degrees of freedom
## Multiple R-squared:  0.2816, Adjusted R-squared:  0.1681
## F-statistic: 2.482 on 3 and 19 DF,  p-value: 0.09204
```

```
2 * 3.023
```

```
## [1] 6.046
```

Adjusted R-squared for our model is 0.1681.Adjusted R2 tells us the percentage of variation explained by only the independent variables that actually affect the dependent variable which is quite less in our case. Here, We can expect 95% of the observed values of y to lie within 2s (2 standard deviations away) of y, i.e.; 6.046

(g) Find and interpret a 95% confidence interval for y when x1 = 20, x2 = 30 and x3 = 25.

```
new_team = data.frame(intrapersonal_range = 20, stress_range = 30, mood_range = 25)
new_team
```

```
##   intrapersonal_range stress_range mood_range
## 1                  20           30         25
```

```
predict(model, newdata = new_team, interval = "confidence", level = 0.95)
```

```
##        fit      lwr      upr
## 1 89.09278 87.18764 90.99792
```

(h) Find and interpret a 95% prediction interval for y when x1 = 20, x2 = 30 and x3 = 25.

```
new_team1 = data.frame(intrapersonal_range = 20, stress_range = 30, mood_range = 25)
new_team1
```

```
##   intrapersonal_range stress_range mood_range
## 1                  20           30         25
```

```
predict(model, newdata = new_team1, interval = "prediction", level = 0.95)
```

```
##        fit      lwr    upr
## 1 89.09278 82.48456 95.701
```