Homework 1 Minh Cao

2025-01-15

Question 1

a. There is a positive association between height of athlete and lenght of jump
b. There is a linear upward trend between two variable => we can use linear regression
c. Choice: 1. Reason: Positive and Upward
d. The response is quantitative, explanatory variable is also quantitative
e. Predited length of the jump is:

```
X = 72
yhat = 6.4285 + 1.0534*X
yhat
```

```
## [1] 82.2733
```

f. The slope of the linear equation
g. No
h. The coach cannot conclude that because this is an observational study, there might be a counfounding variable for the length of the jump (ie, diet, spleep, etc)
i. It will stay the same because we will still calculate the distance length in inches then convert it to feet
j. It will not stay the same because the scale of explanatory variable will change. This would lead change to the intercept
k. The correlation will stay the same. When we scale the y value (ie, convert it from inch to feet), the correlation result stay the same

Question 2

a. Yes, researcher could randomly assign participants to be in different group of amounts of time spent outside.
b. We cannot come to this conclusion because it is from an observational study. There can be confounding factors contribute to the decrease in blood pressure (ie, diet, stress level, etc)
c. Number of miles someone jogs each day could make someone go outside and expose to the nature, jogging also help reduces blood pressure => Confounding variable

Question 3

Quantitative variable: BCPM, GPA, VR, PS, WS, BS, MCAT, Apps Categorical variable: Sex

Question 4

a.

- Response: reasoning scores
- explanatory variable: Sex

b.

- Response: AcceptStatus
- explanatory variable: Sex

c.

- Response: MCAT
- explanatory variable: GPA

d.

- Response: AcceptStatus
- explanatory variable: BCPM

Question 5

```
data = read.csv("C:/Users/caoqu/OneDrive/PC/UCI/STUDY/WINTER_25/DATA_210P_Statistical_Methods_I/data/Me
head(data)
```

```
##   Accept Acceptance Sex BCPM  GPA VR PS WS BS MCAT Apps
## 1      D          0   F 3.59 3.62 11  9  9  9   38    5
## 2      A          1   M 3.75 3.84 12 13  8 12   45    3
## 3      A          1   F 3.24 3.23  9 10  5  9   33   19
## 4      A          1   F 3.74 3.69 12 11  7 10   40    5
## 5      A          1   F 3.53 3.38  9 11  4 11   35   11
## 6      A          1   M 3.59 3.72 10  9  7 10   36    5
```

a.

```
accept_df = data[data$Accept == "A",]
reject_df = data[data$Accept == "D",]

print("Five number summary for GPA for admitted students")
```

```
## [1] "Five number summary for GPA for admitted students"
```

```
summary(accept_df$GPA)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   3.140   3.545   3.715   3.693   3.888   3.970
```

```
print("Five number summary for GPA for rejected students")
```

```
## [1] "Five number summary for GPA for rejected students"
```

```r
summary(reject_df$GPA)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   2.720   3.290   3.380   3.385   3.610   3.770
```

b.

- Mean GPA of accepted group is about 3.693, while mean GPA of rejected group is 3.385.
- Min accepted GPA of admitted student is 3.140 while max GPA of rejected student is 3.770 => GPA is not the only factor

Question 6

a.

```r
model = lm(MCAT~GPA, data = data)

summary(model)
```

```
##
## Call:
## lm(formula = MCAT ~ GPA, data = data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -11.4148  -2.5168  -0.1519   2.6653   8.6616
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)    3.923      6.922   0.567    0.573
## GPA            9.104      1.942   4.688 1.97e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4.088 on 53 degrees of freedom
## Multiple R-squared:  0.2931, Adjusted R-squared:  0.2798
## F-statistic: 21.98 on 1 and 53 DF,  p-value: 1.969e-05
```

- Estimated Regression equation: MCAT_hat = 3.923 + 9.104 * GPA

b.

SSE = Sum (yi - yhat_i)**2 = Sum(Yi - (3.923 + 9.104*Xi)**2)

c.

- SLope: 9.104

Interpretation: For one unit increase in GPA, the MCAT score increase by 9.104 points

d.

- The slope is 3.923

Interpretation: When GPA = 0, the MCAT score is 3.923

This interpretation is not meaningful

e.

```
MCAT_30 = 3.923 + 9.104 * 3.0
MCAT_40 = 3.923 + 9.104 * 4.0

sprintf("Mcat score for 3.0 GPA: %f", MCAT_30)
```

```
## [1] "Mcat score for 3.0 GPA: 31.235000"
```

```
sprintf("Mcat score for 3.0 GPA: %f", MCAT_40)
```

```
## [1] "Mcat score for 3.0 GPA: 40.339000"
```

f. The difference is:

```
9.104*2
```

```
## [1] 18.208
```

g. No, since is an observational study, there are many other factors that could contribute in increase MCAT score other than GPA

Question 7

- Randomized experiments help negate the effect of confounder.