Problem Set 1

Alon Rashty

04/04/2021

Contents

About the Course

Question 1 Because we need to identify the causal factor in order to apply the research results on policymaking, otherwise we could "tackle" the wrong one and not get the results we aimed for. Causal inference is not ML main goal, and instead it is interested in the best predictions.

Question 2 We need to have (or to assume) the error is not correlated with the covariates (omitted variables), so we don't have biased estimates. Also, when we analyze experiments we need to have a good enough randomization between the treatment and control groups.

Question 3

- The linearity assumption means that we are looking for the average treatment effect although it might be heterogeneous.

RStudio

```
library(tidyverse)
library(kableExtra)
```

Question 1

```
iris %>%
  select(contains("Sepal") | Species) %>%
  group_by(Species) %>%
  summarise("Average Sepal Length" = mean(Sepal.Length)) %>%
  kbl(caption = "Table 1", align = 'lc') %>%
  kable_classic(full_width = F, html_font = "Cambria")
```

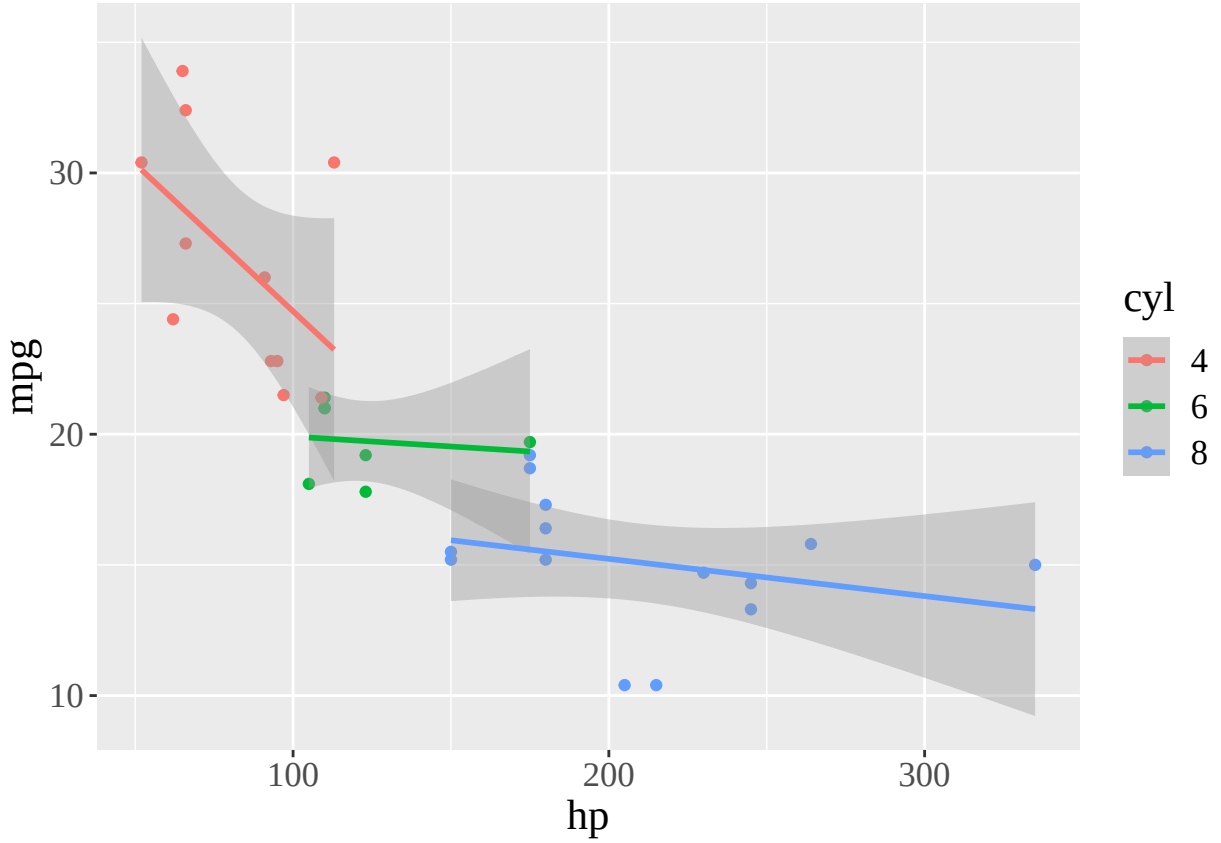
Question 2

```
mtcars %>%
  mutate(cyl = as.factor(cyl)) %>%
  ggplot(aes(x = hp, y = mpg, color = cyl)) +
    geom_point() +
```

Table 1: Table 1

Species	Average Sepal Length
setosa	5.006
versicolor	5.936
virginica	6.588

```
    geom_smooth(method = lm)+
    theme(text=element_text(size=16,  family="serif"))
```



Question 3