

Problem Set 2

Bryson Lyons

2024-09-10

Task 1

Duration in Hours

```
mins_to_hours=function(x){  
  y=x/60  
  return(y)  
}  
  
imdb$duration_hrs = mins_to_hours(imdb$Duration)
```

Unique Directors

There are 918 unique directors in the dataset.

```
length(unique(imdb[["Director"]]))
```

```
## [1] 918
```

Highest Rated Director

After grouping the average rating by director, Roberto Benigni has the highest average rating at 8.6.

```
director_rating<- imdb %>%  
  group_by(Director) %>%  
  summarise(rating = mean(IMDB.Rating)) %>%  
  filter( rating == max(rating))  
  
pander::pander(director_rating)
```

Director	rating
Roberto Benigni	8.6

Unique Ratings

There are 66 unique ratings in the dataset.

```
imdb_rating<- unique(imdb[["IMDB.Rating"]])  
  
length(imdb_rating)
```

```
## [1] 66
```

Rounded Vector

```
rounded<-c()  
  
for(i in imdb_rating){  
  y=round(i)  
  rounded=c(rounded,y)  
}  
  
rounded  
  
## [1] 9 9 8 8 7 8 8 8 8 8 8 9 6 7 7 6 7 8 9 9 8 8 7 7 7 7 6 6 5 6 6 7 6 5 6 5 4 4  
## [39] 6 9 5 9 5 6 4 5 4 6 5 5 4 4 2 5 4 2 3 3 4 2 3 4 2 2 4 2
```

Task 2

According to Angrist and Pischke (2009), a research agenda can be centered around four questions, which they call frequently asked questions (FAQs). These are:

1. What is the causal relationship of interest? 2. What experiment could ideally be used to capture the causal effect of interest? 3. What is your identification strategy? 4. What is your mode of statistical inference?

The first question pertains to the search for cause and effect. Causal relationships allow us to make predictions and tell us what the counterfactual of an event is. The second question posited by the authors directs research design and can help us pick “fruitful research topics”. While the ideal experiment may be hypothetical, it is still worth contemplating, according to the authors. Identification strategies “describe the manner in which a researcher uses observational data to approximate a real experiment (Angrist and Krueger 1999).” Identification strategies can be helpful in pursuing lines of research that may have otherwise been difficult or impossible to analyze. Finally, identifying the mode of statistical inference describes the population being studied, what sample was used, and what assumptions were made. Answering these four questions is requisite for establishing a research strategy in social sciences today, and is something that we, as budding political scientists, should make a habit of.

Task 3

Generating a Random Sample of Test Scores and Assigning Groups

I generated a random sample and created two groups, labeled as “treatment” and “control”. The control group has a mean test score of 70 and the treatment group has a mean test score of 75. Both groups were given a standard deviation of 10.

```

#setting n and means
n=1000
control_mean=70
treatment_mean=75

#set seed and create groups
set.seed(2024)
control<- rnorm(n, mean = control_mean, sd = 10)
treatment<- rnorm(n, mean = treatment_mean, sd = 10)

#true difference in means

true_diff<- treatment_mean - control_mean

treat<- sample(c(rep(1, n/2), rep(0, n/2)), size=n, replace=FALSE)
diff.means<- mean(treatment[treat==1] - mean(control[treat == 0]))

```

Average Treatment Effect

The ATE, or the difference in means between the control and treatment groups, is roughly 5.8. This means that the sample that received the “treatment” scored an average of 5.8 points higher than the control group.

```
diff.means
```

```
## [1] 5.809821
```

Histogram

```

ggplot()+
  geom_histogram(aes(x=control, fill= "Control"), alpha = 0.5) +
  geom_histogram(aes(x=treatment, fill = "Treatment"), alpha = 0.5) +
  labs(x= "Test Scores", y= "Count", fill="Group")+
  theme_classic()+
  geom_vline(xintercept= 70, linetype = "dashed", color="pink", size= 1) +
  geom_vline(xintercept= 75, linetype = "dashed", color="steelblue", size= 1) +
  geom_text(aes(x = 70, y = 25,label = "Control Mean"),size = 2, color = "red") +
  geom_text(aes(x = 75, y = 40,label = "Treatment Mean"), size = 2, color = "blue")

```

