Hypothesis Testing

By: Raashi Maheshwari

Breakdown of DataSet (with R instructions)

```
> summary(Movies2022F.4)
      country       content        imdb_score        Gross          Budget
 USA      :9888   G     : 454   Min.   : 1.070   High  :3597   High  :3322
 UK       :1187   PG    :2378   1st Qu.: 5.810   Low   :4224   Low   :3333
 France   : 353   PG-13:4236   Median : 6.590   Medium:5013   Medium:6179
 Canada   : 291   R     :5766   Mean   : 6.529
 Germany  : 271                3rd Qu.: 7.330
 Australia: 165                Max.   :10.220
 (Other)  : 679
     genre
 Action :2145
 Comedy :3616
 Drama  :1881
 Family :2189
 History:2454
 Sci-Fi : 549

> unique(Movies2022F.4$content)
[1] PG-13 PG    G     R
Levels: G PG PG-13 R
```

summary(Movies2022F.4)

unique(Movies2022F.4$content)

Screenshots from R Studio posted towards the right

Breakdown of DataSet (with R instructions)

```
> unique(Movies2022F.4$content)
[1] PG-13 PG    G     R
Levels: G PG PG-13 R
> unique(Movies2022F.4$country)
 [1] USA                United Kingdom     New Zealand
 [4] Canada             Australia          Belgium
 [7] Germany            China              France
[10] Mexico             Japan              Spain
[13] Hong Kong          Czech Republic     Soviet Union
[16] South Korea        Peru               Italy
[19] Aruba              Denmark            Libya
[22] Ireland            South Africa       Romania
[25] West Germany       Chile              Russia
[28] Netherlands        Hungary            Panama
[31] Greece             Taiwan             Thailand
[34] Slovakia           Bulgaria           Iran
[37] Georgia            India              Norway
[40] Brazil             Bahamas            Sweden
[43] Iceland            Argentina          Colombia
[46] Poland             Israel             Kyrgyzstan
[49] Indonesia          Afghanistan        Switzerland
[52] Dominican Republic
52 Levels: Afghanistan Argentina Aruba Australia Bahamas Belgium ... West Germany
> unique(Movies2022F.4$genre)
[1] Family  Sci-Fi  Action  Drama   Comedy  History
Levels: Action Comedy Drama Family History Sci-Fi
```

unique(Movies2022F.4$content)

unique(Movies2022F.4$country)

unique(Movies2022F.4$genre)

Screenshots from R Studio posted towards the right

Breakdown of DataSet (with R instructions)

```
> unique(Movies2022F.4$Gross)
[1] High   Medium Low
Levels: High Low Medium
> unique(Movies2022F.4$Budget)
[1] High   Low    Medium
Levels: High Low Medium
> unique(Movies2022F.4$genre)
[1] Family  Sci-Fi  Action  Drama   Comedy  History
Levels: Action Comedy Drama Family History Sci-Fi
```

unique(Movies2022F.4$Gross)

unique(Movies2022F.4$Budget)

unique(Movies2022F.4$genre)

Screenshots from R Studio posted towards the right

Breakdown of Dataset (with R Instructions)

```
> tapply(Movies2022F.4$imdb_score, Movies2022F.4$content, mean)
       G        PG     PG-13         R
6.552577 6.301783 6.292623 6.795562
> tapply(Movies2022F.4$imdb_score, Movies2022F.4$Gross, mean)
    High       Low    Medium
6.552438 6.725589 6.347748
> tapply(Movies2022F.4$imdb_score, Movies2022F.4$Budget, mean)
    High       Low    Medium
6.136457 7.087669 6.439676
> tapply(Movies2022F.4$imdb_score, Movies2022F.4$genre, mean)
  Action    Comedy     Drama    Family   History    Sci-Fi
6.500382 6.426980 6.356172 5.908296 7.434874 6.341694
```

```
> tapply(Movies2022F.4$imdb_score, Movies2022F.4$country, mean)
Afghanistan   Argentina       Aruba
   8.430000    8.340000    4.876667
  Australia     Bahamas     Belgium
   6.740909    4.910000    5.806364
     Brazil    Bulgaria      Canada
   8.454615    6.766667    6.408969
      Chile       China    Colombia
   6.970000    6.714839    8.115000
Czech Republic  Denmark Dominican Republic
   6.955000    7.515000    7.470000
     France     Georgia     Germany
   6.847819    5.693333    6.317823
     Greece   Hong Kong     Hungary
   6.817500    6.991818    6.364286
    Iceland       India   Indonesia
   7.940000    7.284000    8.016667
       Iran     Ireland      Israel
   8.127143    7.002759    8.902500
      Italy       Japan   Kyrgyzstan
   7.115152    6.625472    9.750000
      Libya      Mexico Netherlands
   8.525000    7.401290    7.286667
New Zealand      Norway      Panama
   7.535714    7.033846    7.320000
       Peru      Poland     Romania
   5.462000    6.028000    5.007500
     Russia    Slovakia South Africa
   6.285000    6.765000    6.650000
South Korea Soviet Union       Spain
   6.263636    8.830000    7.017952
     Sweden Switzerland      Taiwan
   7.877143    6.113333    8.016667
   Thailand          UK         USA
   6.156875    6.953235    6.440375
West Germany
   7.960000
```

Part A

Alternate Hypothesis #1:

High gross history movies have higher IMDB scores

R Instructions:

```
mean(Movies2022F.4[Movies2022F.4$Budget=='Low' &
Movies2022F.4$genre=='Comedy', ]$imdb_score)
[1] 6.76466

mean(Movies2022F.4[Movies2022F.4$Gross=='High' &
Movies2022F.4$genre=='Comedy', ]$imdb_score)
[1] 6.771587

mean(Movies2022F.4[Movies2022F.4$Gross=='High' &
Movies2022F.4$genre=='Action', ]$imdb_score)
[1] 6.922558

mean(Movies2022F.4[Movies2022F.4$Gross=='High' &
Movies2022F.4$genre=='Drama', ]$imdb_score)
[1] 6.686056

mean(Movies2022F.4[Movies2022F.4$Gross=='High' &
Movies2022F.4$genre=='Family', ]$imdb_score)
[1] 6.090795

mean(Movies2022F.4[Movies2022F.4$Gross=='High' &
Movies2022F.4$genre=='History', ]$imdb_score)
[1] 7.648296

mean(Movies2022F.4[Movies2022F.4$Gross=='High' &
Movies2022F.4$genre=='Sci-Fi', ]$imdb_score)
[1] 6.721778
```

Alternate Hypothesis #2:

High gross movies that are rated R have higher IMDB scores

R Instructions:

```
>
mean(Movies2022F.4[Movies2022F.4$Gross=='High'
& Movies2022F.4$content=='G', ]$imdb_score)

[1] 6.664131

>
mean(Movies2022F.4[Movies2022F.4$Gross=='High'
& Movies2022F.4$content=='PG', ]$imdb_score)

[1] 6.307559

>
mean(Movies2022F.4[Movies2022F.4$Gross=='High'
& Movies2022F.4$content=='PG-13', ]$imdb_score)

[1] 6.432492

>
mean(Movies2022F.4[Movies2022F.4$Gross=='High'
& Movies2022F.4$content=='R', ]$imdb_score)

[1] 7.034664
```

Alternate Hypothesis #3:

High budget movies have a lower IMDB score

R Instructions:

```
> mean(Movies2022F.4[Movies2022F.4$Budget=='Low',]$imdb_score)

[1] 7.087669

> mean(Movies2022F.4[Movies2022F.4$Budget=='Medium',]$imdb_score)

[1] 6.439676

> mean(Movies2022F.4[Movies2022F.4$Budget=='High',]$imdb_score)

[1] 6.136457
```

Part B

Hypothesis #1:

Scenario A

Null Hypothesis:

Movies made in the USA and Australia have the same average imdB ratings

Alternate Hypothesis:

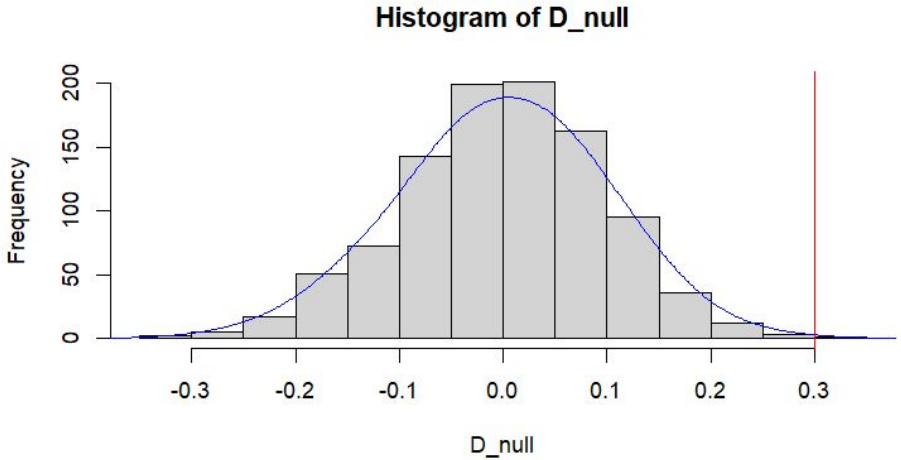
Movies made in Australia tend to have a higher average imdB rating than movies made in the USA

R Instructions:

```
> mean(Movies2022F.4[Movies2022F.4$country == 'USA',]$imdb_score)

[1] 6.440375

>

> mean(Movies2022F.4[Movies2022F.4$country ==
'Australia',]$imdb_score)

[1] 6.740909

> PermutationTestSecond::Permutation(Movies2022F.4,
"country","imdb_score", 1000, "USA", "Australia")

[1] 0.001
```

Scenario A

Conclusion



Since the p-value is lower than 0.05 (p-value = 0.001), we have evidence against the null hypothesis, hence we reject the null hypothesis

We can say that movies made in Australia and the movies made in the USA do not tend to have the same imdb averages

Hypothesis #2

Scenario B

Null Hypothesis:

Movies made in Chile and Denmark have the same average imdB rating

Alternate Hypothesis:

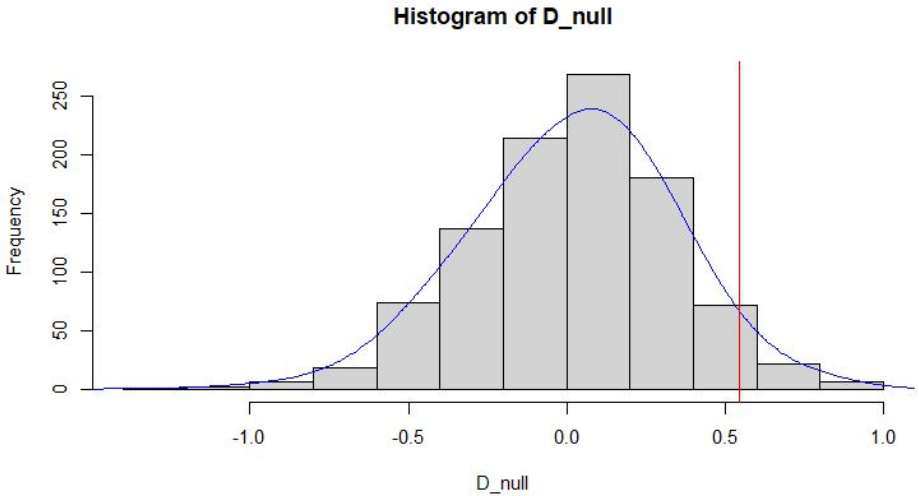
Movies made in Denmark have a higher average imdB rating than movies made in Chile

R Instructions:

```
> mean(Movies2022F.4[Movies2022F.4$country ==
'Chile',]$imdb_score)

[1] 6.97

> mean(Movies2022F.4[Movies2022F.4$country ==
'Denmark',]$imdb_score)

[1] 7.515

> PermutationTestSecond::Permutation(Movies2022F.4,
"country","imdb_score", 1000, "Chile", "Denmark")

[1] 0.037
```

Scenario B

Conclusion



Since the p-value is 0.037, we reject the null hypothesis. We have enough evidence to say that Movies made in Chile and Denmark do not have the same average imdB rating

Hypothesis #3

Scenario C

Null Hypothesis:

PG-13 rated movies have the same imdB score as PG rated movies

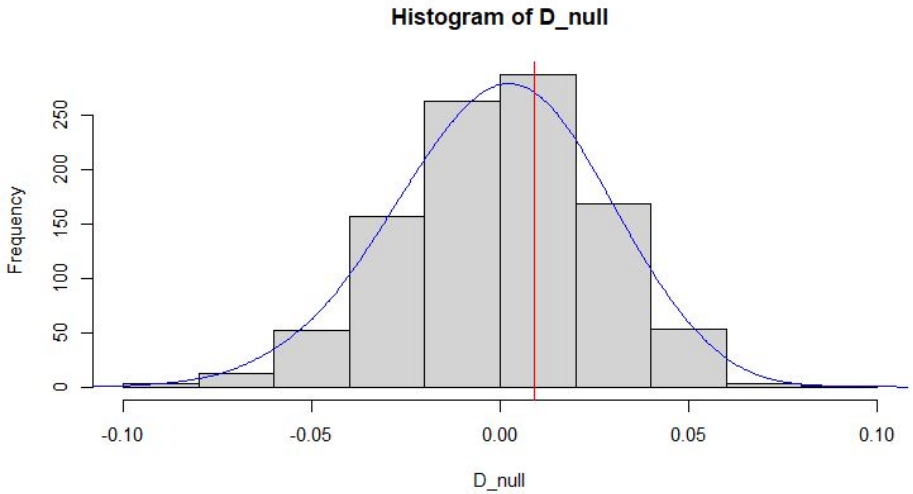
Alternate Hypothesis:

PG-13 movies tend to have a higher imdB score than PG rated movies

```
mean(Movies2022F.4[Movies2022F.4$content=='PG',
]$imdb_score)

[1] 6.301783

> mean(Movies2022F.4[Movies2022F.4$content ==
'PG-13',]$imdb_score)
[1] 6.292623

PermutationTestSecond::Permutation(Movies2022F.4,
"content","imdb_score", 1000, "PG-13", "PG")
[1] 0.369
```

Scenario C

Conclusion



We fail to reject the null hypothesis as the p-value from permutation test is 0.369, which is more than 5% hence there is strong evidence that the null hypothesis may be true.