Lab Report 2: Hypothesis Testing

Women's Representation in the World

DUE: October 14 at 5 PM

You work at the United Nations and have been tasked with improving women's representation in governments around the world. Your boss gives you a dataset including data on 170 countries in the world in 2002. Your boss is particularly interested in understanding regional disparities in women's representation as well as whether changes in electoral systems might improve women's representation. More specifically, she asks you to 1) determine whether there are regional differences in women's legislative representation and 2) whether women's representation is different in proportional representation systems compared to other systems.

You can read in the data "oilwomen.csv" with the following command (or download and import it from CANVAS):

```
dataset <- read.csv("https://raw.githubusercontent.com/ilaydaonder/POLS209/Lab-Report-2/oilwomen.csv")
```

Your boss is interested in the following regions: (a) the Middle East/North Africa (me_nafr) (b) Sub-Saharan Africa (ssafrica) (c) Asia (asia) (d) Latin America (latin). Each of these is coded with a 1 if the country is in that region and a 0 otherwise. For example, Colombia is coded with a 1 for "latin" and a 0 for all other regional variables.

Three other variables are important for your analysis:

- **female_seats**: The percentage of seats in the legislature held by women
- **female_seats_15**: Coded with a 1 if the percentage of seats held by women in the legislature is greater than 15% (the mean)
- **PR**: A binary variable indicating whether the country has a proportional representation system. It is coded 1 if the country has a proportional representation system.

To answer your boss's question, first test the null hypotheses that the percentage of women in the legislature does not differ in each region from the rest of the world (all other countries not in that region). Report your results in the following table. Bold the p-values that are statistically significant.

Region	**Region Average**	**Rest of World Average**	**Test Statistic**	**P-Value**
Middle East/North Africa				
Sub-Saharan Africa				
Asia				
Latin America				

Then subset the data to create a separate dataset for each region. Test the hypothesis that the average percentage of women in the legislature in a pair of regions is equal. In other words, conduct a difference of means test for each combination of regions. Report your results in the following table. Bold the p-values that are statistically significant.

Region 1	**Region 2**	**Region 1 Mean**	**Region 2 Mean**	**P-Value**
Middle East/North Africa	Sub-Saharan Africa			
Middle East/North Africa	Asia			
Middle East/North Africa	Latin America			
Sub-Saharan Africa	Asia			
Sub-Saharan Africa	Latin America			
Asia	Latin America			

Finally, test the hypothesis that proportional representation systems and being above average in terms of women's representation (female_seats_15) are independent. Report a crosstabulation table and whether you can reject the null based on the test statistic and p-value.

Summarize your results (and include the tables) in a 1-2-page memo. In the memo, interpret the sample means and discuss differences across regions. Do any particular regions tend to be associated with greater women's representation? In what region(s) should the United States focus its efforts at increasing women's representation? Do proportional representation systems seem to improve women's representation? Submit the memo (which should include the tables) and copy/paste your R code at the end.