Michael Hawkinson

STAT 301-01

4-23-24

**Analysis Project**

**Introduction**

I chose to perform an analysis using the country datasets (United States, Mexico, Canada, and Japan). The provided datasets have rows of data that represent the total trade value between two countries in a given month. Each row has the following variables:

* Year: what year the transactions took place (2021 in each row of the datasets)
* Period: the month the transactions took place, in yyyymm format
* Period Desc.: the date the transactions took place, in yyyy-mm-dd format (although there is no granularity on day, as all data is shown as occurring on the first of the month)
* Trade Flow Code: A numeric value, showing the type of trade:
  + 1 – Imports
  + 2 – Exports
  + 3 – Re-exports (exporting after previously importing)
  + 4 – Re-imports (importing after previously exporting)
* Trade Flow: The text value associated with the Trade Flow Code
* Reporter: The country reporting the transaction (the origin country/territory)
* Partner Code: A numeric value designating the destination country/territory
* Parter: The text value associated with the Partner Code
* Commodity: The commodity type (All Commodities in each row of the datasets)
* Trade Value (US$): The total value of the transactions in USD.

I chose to research the total dollar value of exported goods across each country and see if there are any significant differences between countries. To do this, we will compare the means of total trade value of exports using the following hypotheses:

**Methods**

Given that each of the countries are independent of each other, there is one independent variable (Reporter) and one numeric independent variable (Trade Value (US$)), one-way ANOVA can be used to test for a significant difference in means.

The data was processed in R, filtering rows from each dataset that correspond to exports and placed into a merged data frame. Additionally, each dataset had multiple rows where the Partner was listed as “World”: these appeared to be total values for the given month, so those values were also removed.

**Conclusion**

The one-way ANOVA results in a test statistic (F-value) of 20.03, with a p-value of 0. Because of this, the null hypothesis can be rejected. We can say with 95% confidence that at least 2 means are significantly different.

To expand on this, Tukey’s test can be used to see where the significant differences are.

Tukey’s multiple comparison of means, 95% confidence

`Trade Value (US$)` ~ Reporter

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **$Reporter** | **diff** | **lwr** | **upr** | **p adj** |
| Japan-Canada | 96535945 | -70950495 | 264022384 | 0.4490352 |
| Mexico-Canada | 178238560 | -31271573 | 387748694 | 0.1271591 |
| USA-Canada | 471459466 | 305365898 | 637553035 | **0.0000000** |
| Mexico-Japan | 81702616 | -126855575 | 290260807 | 0.7455545 |
| USA-Japan | 374923521 | 210032357 | 539814686 | **0.0000000** |
| USA-Mexico | 293220906 | 85779625 | 500662187 | **0.0016103** |

Based on the results of Tukey’s test, the United States’ mean export trade value differs significantly from the other three countries.

I was surprised to see that the United States had large differences in exports compared to the other three countries. It is well known that the US is one of the top importers in the world, but I was not aware of the US being a top exporter prior to performing this analysis. Generating the mean trade value for each country makes it clear that the US exports much more than the rest:

A graph showing the value of trade

Description automatically generated

One limitation of this analysis is that this data is for a single year. A nation’s economy can differ drastically from year to year, so it would be best to take data from a decade (or longer) to get a more holistic representation of trade values. Additionally, adding more countries for comparison would result in a better analysis: it would be especially interesting to compare the US export values to a country like China, who exports more than any other country in the world.

Github link: https://github.com/m-hawkinson/stat\_301\_analysis\_project