Summer Olympics- Sample Paper

Prof V

Introduction

Winning big at the Olympics brings national prestige and an increased presence on the world stage for a country. This leads to a country to increase its image and popularity, in turn could cause increases in tourism a potential bid to host the Olympics, and other strong political relationship. Therefore, it may be useful to predict how many medals a country would win. What factors might impact a country's success?

Filipino gymnast, Carlos Yulo, made headlines as the first male to win an Olympic gold medal in the history of the Philippines. But he continued to grow in popularity as his earnings from the country caught attention. He earned a cash prize of 10 million Philippine pesos ($172,519 USD), a brand new condo valued at over $400,000 USD, a lifetime supply of ramen and other incentives. In comparison, the US awards its gold medalists a "measly" $38,000 for winning the gold. While governments in Great Britain and Sweden do not offer any direct cash incentives. In this unit we will explore data from the 2024 Paris Olympics, how cash incentives and national factors impact medal counts for each country. We will explore the countries with the top 24 medal counts.

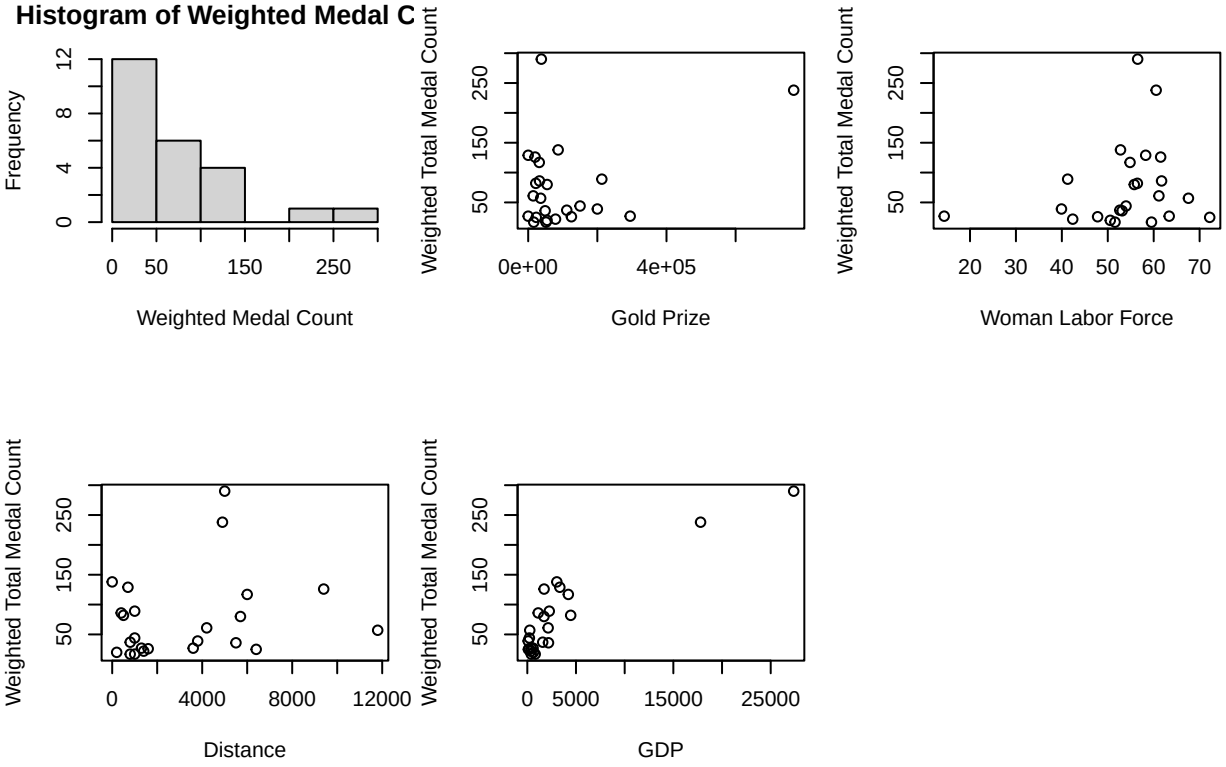
- Do countries with higher gold cash incentives have a higher weighted medal counts?
- Do countries with higher female labor force have a higher weighted medal counts?
- Do countries that are furhter from Paris have a lower weighted medal counts?

Data Summary

Data Sources

The data were compiled from several sources including ESPN, Olympic International Committee, and World Bank. Some values were cross references on Wikipedia. We have no reason to question the credibility of these sources.

Exploratory Data Analysis



EDA Summary

Our research variables Gold Prize, Woman Labor Force, and Distance did not have a strong linear relationships with correlations of 0.352, 0.209, and 0.217 respectively. However, GDP has a surprisingly strong relationship with a correlation of 0.987. Additionally, multicollienrty was a concern with several variables with high VIFs. After performing stepwise with entry and stay significance of 0.20, the variables GDP, YOSilver, and Woman Labor Force were determined to be an important set of predictors. Therefore, we will begin model building with those variables.

Methods and Analysis

After EDA, we fit the model with GDP, YOSilver, and Woman Labor Force. With a pvalue less than 0.05, the overall model was significant. Then we evaluated "Woman Labor Force" with a t test. With a pvalue of 0.057, this does not meet our significance level. However there is moderate evidence this is an important predictor so we will keep it in the model.

... continue assessing and adding qual, int...

... residual analysis ...

Results

Our final model is .... the ways we assess it are statistically....

Conclusions

Practically our model is .... the way we use it...

Appendix A: Data Dictionary

Variable Name	Abbreviated Name	Description
Country	Country	The team name as identified by the country.
WeightedTotal	Weighted Total	Weighted medal count from Paris 2024 (4 points for each gold, 2 for each silver, 1 for bronze)
WomanLaborForce	Woman Labor Force	Percentage of woman working in the country in 2023

Appendix B: Data Rows

	Country	WeightedTotal	Total	TGold	TSilver	TBronze	GoldPrize	SilverPrize
1	Australia	126	53	18	19	16	20000	15000
2	Belgium	20	10	3	1	6	54606	32764
3	Brazil	36	20	3	7	10	49000	29000
4	Canada	61	27	9	7	11	14572	10929
5	China	238	91	40	27	24	768000	384000
6	Denmark	17	9	2	2	5	15962	11971

	BronzePrize	YOGold	YOSilver	YOBronze	YOTotal	GDP	Population
1	10000	4	8	4	16	1723.8	26.71
2	21843	2	3	2	7	632.2	11.74
3	20000	2	4	7	13	2173.7	212.00
4	7286	0	3	6	9	2140.1	39.74
5	192000	18	9	9	36	17794.8	1419.32
6	7981	2	1	1	4	404.2	5.98

	WomanLaborForce	TotalAthletes	Distance
1	61.52	460	9400
2	50.55	165	200
3	53.13	277	5500
4	61.14	316	4200
5	60.54	288	4900
6	59.52	123	800

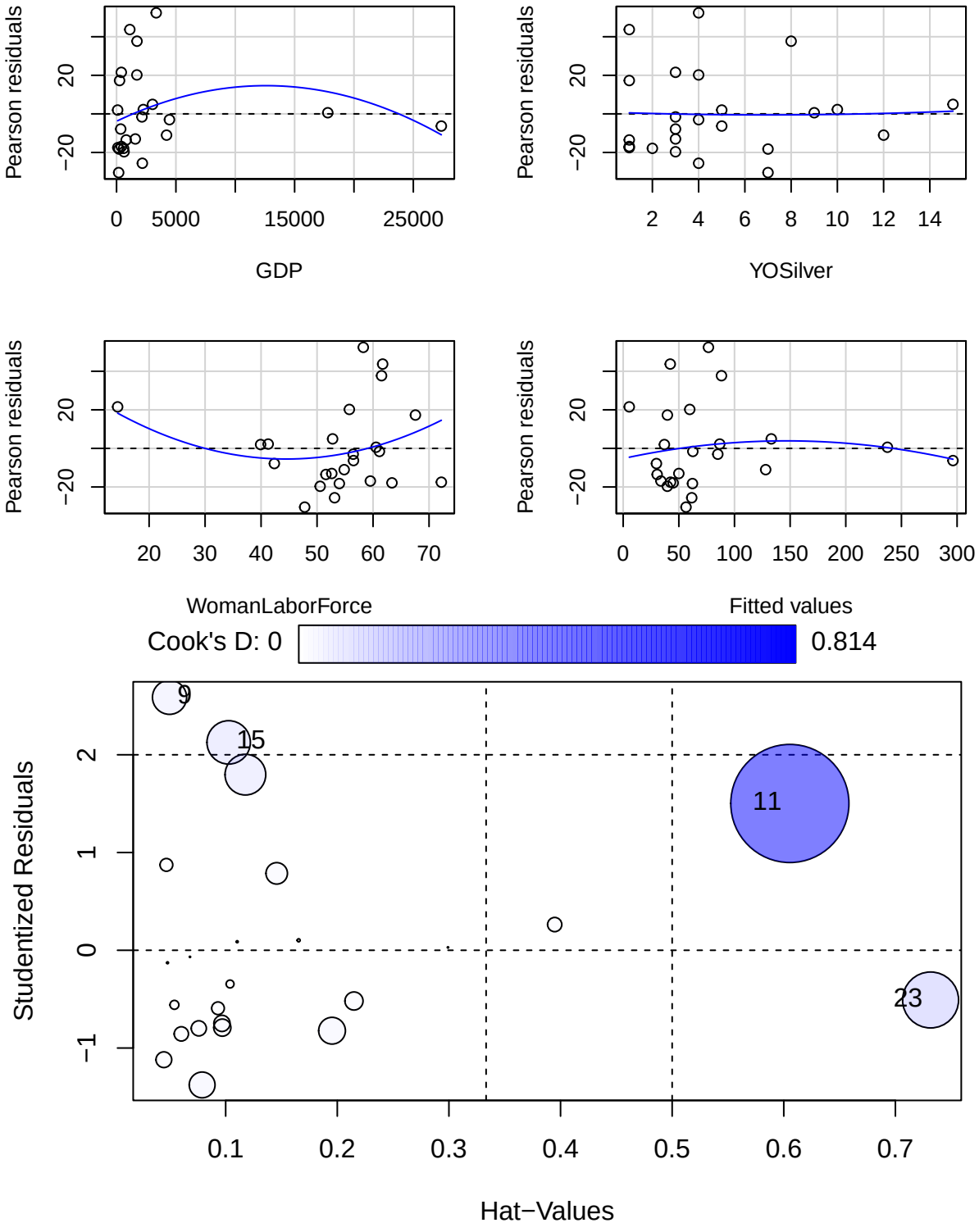
Appendix C: Final Model Output and Plots

```
Call:
lm(formula = WeightedTotal ~ GDP + YOSilver + WomanLaborForce,
    data = olympics2024)

Residuals:
    Min      1Q  Median      3Q     Max
-30.442 -17.069  -4.666   8.020  52.389

Coefficients:
                  Estimate Std. Error t value Pr(>|t|)
(Intercept)     -2.839e+01  2.577e+01  -1.102 0.283637
GDP              8.969e-03  8.191e-04  10.949 6.74e-10 ***
YOSilver         5.827e+00  1.379e+00   4.225 0.000416 ***
WomanLaborForce  8.881e-01  4.412e-01   2.013 0.057799 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 23.54 on 20 degrees of freedom
Multiple R-squared:  0.9013,    Adjusted R-squared:  0.8864
F-statistic: 60.85 on 3 and 20 DF,  p-value: 3.112e-10
```



	StudRes	Hat	CookD
9	2.5880835	0.04969438	0.06815038
11	1.5017881	0.60546004	0.81416644
15	2.1267190	0.10267508	0.11000542

```
23 -0.5084191 0.73153058 0.18286442
```

Appendix D: References

Background

Data Sources

Additional Help