Script:

<https://github.com/gregorioCPcG/sula_liberta>

Desde 2017, os times que disputam Copa Libertadores no mesmo ano\* jogam a Copa Sulamericana. O que se quer saber aqui é quem se sai melhor: os times que disputam a segunda competição do continente desde o início ou esses oriundos da liberta que entram em fases mais adiantadas da competição?

Em 2021, os 4 semifinalistas não vieram da liberta. Esse fato me chamou à atenção, já que supostamente eles teriam se saído melhor na temporada anterior em seus países, dado que se classificaram à principal competição do continente. Mas será que 2021 foi uma exceção? É o que pretendo esclarecer (se os dados permitirem) nas próximas linhas.

Para fins de classificação sobre resultado final é designado uma pontuação para cada etapa atingida por equipe:

* Segunda fase – 0 (entre 2017 e 2020, os times da Libertadores entravam nessa etapa)
* Oitavas de Finais – 1 (Com a adoção da fase de grupos na Sulamericana, os terceiros da liberta passaram a entrar nessa etapa)
* Quartas de finais – 2
* Semifinais – 3
* Vice – 4
* Campeão – 5

A hipótese a ser testada é que é origem dos times (duas opções aqui Libertadores ou fases anteriores) teria correlação com o resultado final na competição.

A pontuação não é cumulativa, a título de exemplo Red Bull Bragantino, atual vice, obteve 4 pontos nesse sistema de pontuação. Já o campeão Athletico obteve 5. Não será dado peso diferente para os anos, já que nosso objetivo não é comparar equipes e sim competições.

*O resultado: rodamos um teste chamado regressão linear e não encontramos correlação.*

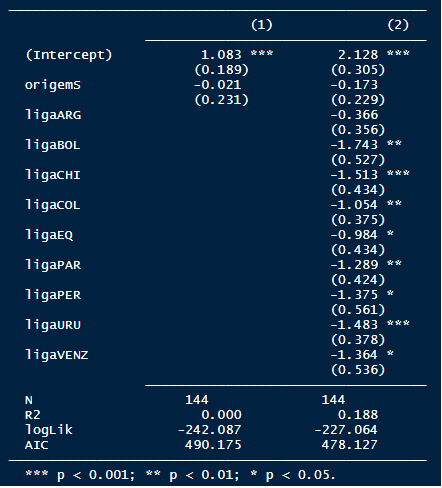
A força da Liga?

É possível ter um cenário mais amplo adicionando uma variável para verificar se há outro fator correlacionado.

A força da liga foi então testada, comparamos as nove ligas da América do Sul com a do Brasil e descobrimos que é esse fator se correlaciona com o desempenho na Sulamericana! Com exceção da liga argentina (em que não há correlação e, portanto nada pode ser dito), todas as outras apresentaram correlação negativa. Ou seja desempenho inferior de 8 ligar em comparação aos times oriundos do brasileirão.

A conhecida dominância dos times brasileiros na libertadores, também tem se replicado na segunda competição da América do Sul.

Veja abaixo os resultados:



Na primeira coluna consta o modelo sem a variável “País” e não há correlação encontrada.Já na segunda há alguns casos com significância estatística (a presença de asterisco).

Um time da Bolívia tende a ter -1,74 pontos que um time brasileiro. Um time chileno, por sua vez, -1,513; e assim sucessivamente. No caso da Argentina por não ter asterisco, não há significância estatística. Já a origem do time (se é libertadores ou fases anteriores) não é significativo.

A comparação entre os dois modelos também pode ser feita mediante o valor de R2, ao passo que o primeiro tem 0.188 e o segundo é zero isso denota a significância encontrada na variável país. Já o AIC quanto menor, melhor o ajuste e a robustez e nesse caso o segundo modelo também se sai melhor. A imagem abaixo apresenta graficamente esses achados

