

Hemuppgift 2

R för samhällsvetare. Stockholms universitet. Sommar 2020.

Besvaras senast måndag 29 juni kl 17.

1. Inlämningsuppgifter besvaras genom att ladda upp ett R-skript på Athena . Om du föredrar går det även bra att redovisa dina svar i ett worddokument, pdf, powerpoint eller dylikt. Döp filen till "*Hemuppgift_2.R*". Du behöver inte ange ditt namn.
2. Det väsentliga är inte att dina resultat är snyggt presenterade. Huvudsaken är att du finner någon form av svar på frågan. Oavsett filformat ska det tydligt framgå vilka kommandon du använder för att lösa respektive uppgift. Skriv kommentarer i din kod (#).
3. Om du kör fast på någon fråga – lämna inte blankt. Gör så gott du kan och förklara kort vad som inte fungerar.

Uppgift

Använd datafilen *hu2_lifeexp.RData*. Filen innehåller data per land för följande variabler:

1. `country` = land
2. `continent` = världsdel som detta land tillhör.
3. `child_mort` = barnadödlighet, promille av barn <5 år.
4. `econ_freedom` = ett index över ekonomisk frihet.
5. `gdp` = bruttonationalprodukt (BNP på svenska).
6. `gini` = ginkoefficienten för individers inkomst, hur stor andel av detta lands inkomster som måste omfördelas för att alla individer ska ha samma inkomst.
7. `hdi` = Human Development Index, från FN.
8. `health_exp` = andel av detta lands utgifter för hälso- och sjukvård som finansieras via skatter.
9. `life_expectancy` = förväntad medellivslängd.
10. `women_econ_op` = index över kvinnors ekonomiska möjligheter.

Källa: Ourworldindata.org, Världsbanken, Fraser Institute, GapMinder.

Lös följande:

1. Ta fram median och medelvärde för alla variabler som innehåller siffror.
2. Skapa ett histogram var för de variabler som innehåller siffror, totalt 8 stycken.
Tips: Om du inte vill skriva ett separat kommando per diagram (du får göra så om du vill), kan du titta närmare på något av följande kommandon: `pivot_longer()`, `facet_wrap()`, `map()`, `for()`.
3. Skapa punktdiagram för följande variabler (välj x och y själv):
 - a. `life_expectancy` och `gdp`
 - b. `life_expectancy` och `gini`
 - c. `gdp` och `gini`

4. Gör samma sak som i fråga 3 men lägg till den rätta linjen, beräknad med hjälp av minsta kvadratmetoden för diagrammen (välj x och y själv). Tips: `geom_smooth()`.
5. Ta de tre kombinationerna i fråga tre och beräkna linjära regressionsmodeller för dem, i stil med $y = a + b \cdot x$, där x och y är variabler som vi har data på och a och b är parametrar som ska beräknas. Du får själv välja vilken variabel du sätter som x och y .
 - a. Notera: I R skrivs modeller av denna typ som $y \sim x$. De beräknas med `lm()`-kommandot. Du behöver inte presentera dina resultat på något särskilt sätt.
6. Nu ska du beräkna en samling regressioner. `life_expectancy` är den beroende variabeln (alltså y) i alla regressioner.
 - a. Beräkna en modell för varje annan variabel i tabellen som innehåller siffror. Tips: `map()`.
 - b. Redovisa resultaten från 6a i en stor tabell. Till exempel med hjälp av kommandot `stargazer()`.