

I. CO2-utslipp (teller 40%)

I denne oppgaven skal vi se på sammenhengen mellom lands CO2-utslipp og BNP. Vi skal bruke data for CO2-utslipp i tonn per person i 2019 fra *Our world in data* og sette sammen med tall på BNP per person og befolkning fra World Development Indicators.

Komma- og tabulatorseparerte filer er vedlagt denne oppgaven for nedlasting og du kan også hente dataene fra nettet med kommandoen

```
read.csv('https://www.uio.no/studier/emner/sv/oekonomi/ECON2130/v21/timeplan/data/co2.csv')
```

I datasettet er det følgende variable

<code>iso_code</code>	Landkode
<code>land</code>	Landnavn
<code>co2_per_capita</code>	CO2-utslipp per person (2019)
<code>bnp_per_capita</code>	BNP per person (2019)
<code>befolkning</code>	Antall innbyggere (2019)
<code>oecd</code>	Landet var medlem i OECD i 2019

- Last dataene inn i en dataramme i R. Lag et histogram over variabelen `co2_per_capita`. Kommenter fordelingen til variabelen.
- Lag et spredningsdiagram (scatter plot) med BNP per person på x-aksen og CO2-utslipp per person på y-aksen. Diskuter hva vi kan lese ut av diagrammet.

Vi antar at landene i datasettet er et tilfeldig utvalg av alle verdens land.

- Estimer forventet CO2-utslipp per capita i verdens land. Er estimatoren din forventningsrett og/eller konsistent, og hvilken fordeling har den?
- Lag et 90 % konfidensintervall på estimatet fra oppgave c). Vær klar på hvilke forutsetninger du gjør og hvordan du vil tolke intervallet.

OECD-landene er ofte sett på som en gruppe av rike land. Vi ønsker å teste om CO2-utslippene er høyere i OECD-landene enn i de andre landene.

- Sett opp de relevante hypotesene for å teste dette.
- Gjennomfør en t-test av hypotesen at CO2-utslipp per capita er høyere i OECD-land enn i andre land med et signifikansnivå på 1 %. Vær klar på hvilke forutsetninger du gjør og hvordan du vil tolke utfallet av testen.

Colombia ble medlem av OECD 28. april 2020.

- Er det riktig å konkludere fra det du fant på oppgave f) at vi kan forvente at CO2-utslippene i Colombia da går opp?
- Vi ønsker å gjennomføre testen fra oppgave f) med å bruke medianer i stedet for gjennomsnitt. Drøft om dette er en bedre eller dårligere tilnærming. Finn median CO2-utslipp i OECD-landene og de andre landene.

For å teste signifikansen til testen basert på median ønsker vi å bruke simulering. Spesifikt vil vi 1000 ganger definere 36 tilfeldige land som «OECD-land» og finne medianen til denne gruppa.

- Lag et R-skript som gjennomfører denne simuleringen og finn forkastningsverdien for medianen i en test med signifikansnivå 5 %. Gjennomfør testen.

II. Evaluering av emne (teller 30%)

I emnet ECON2130 hadde vi en midveiseevaluering hvor 22 studenter svarte på noen spørsmål om lærerne, undervisningen, og sin egen opplevelse av kurset. Blant spørsmålene de svarte på var om de deltok på spørsmål og svar (QnA) rundene alltid, noen ganger eller aldri. De svarte også på om de hadde problemer med å få utbytte av forelesningene eller seminarer på grunn av mangelfulle forkunnskaper. I tabellen nedenfor kan du se hvordan svarene fordelte seg

Deltok på QnA	Mangelfullt utbytte på grunn av mangelfulle forkunnskaper			
	Ja	Til en viss grad	Nei	Total
Aldri	0	3	2	5
Noen ganger	1	5	7	13
Alltid	1	1	2	4
Total	2	9	11	22

Vi lager en ny variabel, MangelfulltUtbytte som er 1 om de svarer «Ja», 0 om de svarer «Til en viss grad» og -1 om de svarer «Nei».

- Hva er forventet verdi av variabelen MangelfulltUtbytte for en tilfeldig trukket respondent.
- Hva er forventet verdi av variabelen MangelfulltUtbytte for en tilfeldig respondent, gitt at vedkommende aldri deltok på QnA?

Foreleserne lurer på om tilbøyeligheten til å delta på QnA forelesninger er ulik for de med mangelfulle forkunnskaper og de som ikke har mangelfulle forkunnskaper.

- Formuler en hypotese og en test du ville bruke for å svare på dette. (Du skal ikke gjennomføre testen!)

Det var totalt 151 studenter oppmeldt til eksamen i kurset, og 22 som valgte å svare på spørreundersøkelsen.

- Diskuter om dette er et tilfeldig utvalg og i hvilken grad dette har betydning for konklusjonene vi kan trekke fra dette datamaterialet.

III. Valg av estimator (teller 20%)

La X og Y være to uavhengige stokastiske variabler. X har forventning μ og varians σ^2 , mens Y har forventning 2μ og varians $2\sigma^2$. La $Z = aX + bY$ for to konstanter a og b .

- Finn forventning og varians til Z .
- Vi ønsker å bruke Z til å estimere μ . Hvordan vil du velge a slik at Z blir en forventningsrett estimator av μ for hver av tilfellene $b = \frac{1}{2}$; $b = \frac{1}{3}$ og $b = \frac{1}{4}$?
- Hva blir variansen til de tre forventningsrette estimatorene du fant i b)? Hvilken estimator vil du foretrekke? Forklar intuisjonen bak hvorfor dette er den beste estimatoren.

IV. Betingede sannsynligheter (teller 10%)

Vi kaster to rettferdige terninger der alle utfall fra 1 til 6 er like sannsynlige. La T_1 være antallet øyne opp på den første terningen, mens T_2 er antallet øyne som vender opp på den andre terningen. Endelig er $S = T_1 + T_2$ summen av antall øyne.

- a) Hva er sannsynligheten $P(S \geq 5 | T_1 = 3)$?
- b) Hva er sannsynligheten $P(S \geq 5 | T_1 \leq 2)$?