

Statistikk og statistisk programmering, våren 2022

Obligatorisk oppgave 4

Innleveringsfrist: Fredag 11. mars 2022, kl. 23.59

Leveres på Canvas.

De Python-baserte oppgavene må gjøres for å få godkjent innleveringen.

Du må levere selve Python-fila, og ikke bare bilde av koden. I tillegg skal du levere en pdf-fil med svar på resterende oppgaver. Du kan også skrive Python-koden i Word så vi selv kan kopiere det over i egen fil. Om du skriver oppgavene i Word, eller om du skriver for hånd og limer inn bilde av dette i Word-fila, spiller ingen rolle. Det som er viktig er at det er oversiktlig og lesbart for oss! Outputen du får når du kjører fila tar du et bilde av og limer inn i pdf-en.

Oppgave 1

I denne oppgaven skal du bruke Python i spørsmål b og c.

- a) Anta at du kaster en terning. Vi kaller antall øyne som terningen viser for X . Finn forventningsverdien og standardavviket til X .
- b) Lag et Python-program som simulerer flere omganger med terningkast.

Programmet skal be brukeren taste inn hvor mange terningkast som skal gjøres i hver omgang og hvor mange omganger som skal gjøres.

Programmet skal så simulere disse terningkastene og summere hvor mange øyne terningen viser i hver omgang, og skal også regne ut gjennomsnittet av antall øyne i hver omgang.

Programmet skal så tegne et histogram med 30 søyler som viser disse gjennomsnittene.

Kjør programmet med en del ulike tall, både små tall, store tall og veldig store tall.

Tips:

For å generere tilfeldige heltall mellom 1 og 6, kan man bruke

```
random.randint(1, 6)
```

men da må man først importere pakken `random` fra Pythons standardbibliotek ved å skrive

```
import random
```

Når man skal holde styr på totalt antall øyne i de ulike omgangene, kan man for eksempel definere et array med N elementer initialisert med bare 0-er ved hjelp av `numpy.zeros(N)`

- c) Utvid programmet i spørsmål b slik at det tar alle gjennomsnittene som du beregner, og finner og skriver ut gjennomsnittet av disse og utvalgsstandardavviket for disse.

Legg merke til at utvalgsstandardavviket som beregnes av programmet blir i nærheten av $\frac{\sigma}{\sqrt{n}}$ hvor σ er standardavviket du fant i spørsmål a, og n er antall kast i hver omgang.

Oppgave 2

Denne oppgaven skal løses ved hjelp av Python.

På den greske øya Kripos kryr det av villkatter.

Myndighetene har plukket ut noen katter som de av ulike grunner har undersøkt og blant annet funnet vekten av kattene og vekten av deres hjerter. Resultatet av målingene finner du i følgende fil:

http://www.it.hiof.no/statok/katters_vekt.csv

De fire feltene i filen er: løpenummer, kjønn, kattens vekt i kg og hjertets vekt i gram (dette er altså en fil jeg har funnet på nettet med faktiske målinger av vekten til katter – ikke tenk for mye på hvordan de har klart å veie hjertet ...).

Vi ønsker å sjekke om det er rimelig å anta at vekten av kattene og vekten av deres hjerter begge er normalfordelte.

Lag et Python-program som leser inn data fra fila og lager et normaltestplott (QQ-plott) først for kattens vekt og så for hjertens vekt.

Tips:

For å lese inn data fra fil, kan man bruke `numpy.genfromtxt`.

For å lage et normaltestplott kan man gjøre følgende import:

```
from statsmodels.graphics.gofplots import qqplot
```

og så bruke `qqplot` på sine data.

Oppgave 3

I denne oppgaven skal du ikke bruke Python.

Bruddstyrken til tau kan måles i hvor mange kilo belastning det tåler før det ryker. Vi kaller bruddstyrken for X , og antar at $X \sim N(\mu, \sigma)$.

Vi måler 10 tau og finner følgende bruddstyrke for disse:

1525, 1401, 1593, 1544, 1448, 1519, 1487, 1491, 1446, 1546

- a) Finn basert på disse 10 målingene, et punktestimat for μ .
- b) Finn basert på disse 10 målingene, et 95 % konfidensintervall for forventningsverdien til bruddstyrken dersom vi antar at $\sigma = 57$. Avrund grensene til nærmeste hele kilo.
- c) Finn lengden av konfidensintervallet.
- d) Hvor mange tau må måles for at lengden av konfidensintervallet ikke skal overstige 50?

Vi får inn fire andre tau hvor vi ikke kjenner standardavviket til bruddstyrken. Vi måler bruddstyrken til disse, og finner følgende:
1200, 1250, 1150, 1300

- e) Finn et 95 % konfidensintervall for forventningsverdien til bruddstyrken til disse tauene basert på disse fire målingene. Også her skal du avrunde grensene til nærmeste hele kilo.

Oppgave 4

I denne oppgaven skal du ikke bruke Python.

Statistisk sentralbyrå utførte en meningsmåling blant 600 norske jenter i alderen 15–19 år, der spørsmålet var om de kunne tenke seg å ta ingeniørutdanning. Av alle norske jenter i alderen 15–19 år, vil en andel svare ja. Vi kaller denne andelen p .

60 av de 600 jentene svarte ja på spørsmålet.

Finn et 95 % konfidensintervall for p basert på denne spørreundersøkelsen.

Oppgave 5

Denne oppgaven skal løses ved hjelp av Python.

Lag et Python-program som beregner et konfidensintervall for sannsynligheten p i en binomisk fordeling. Programmet skal fungere slik:

Programmet skal be brukeren taste inn utvalgets størrelse og andelen av utvalget som er i favør av det man skal sjekke. (Hvis du skulle brukt dette programmet på forrige oppgave, ville disse tallene altså vært henholdsvis 600 og 60). Programmet skal også be brukeren angi konfidensintervallet **i prosent**.

Basert på disse tallene som brukeren har tastet inn, skal programmet sjekke om tallene oppfyller kravene til en tilnærmet normalfordeling (se regel 6.12 side 252 i boka), og skrive ut en beskjed om resultatet av denne sjekken.

Uansett om tilnærmingen er god eller ikke, skal programmet regne ut et konfidensintervall for p og skrive ut dette med en passende ledetekst.

Sjekk at programmet virker som det skal ved å kjøre det med de tallene du fant i forrige oppgave.

Oppgave 6

Denne oppgaven skal løses ved hjelp av Python.

Lag et Python-program som beregner et punktestimat og et konfidensintervall for standardavviket σ til målinger som er hentet fra en normalfordeling.

Målingene kan du legge i en kommaseparert fil som leses inn av programmet.

Programmet skal be brukeren taste inn et konfidensintervall i prosent.

Basert på dette skal programmet regne ut et punktestimat og et konfidensintervall for σ , og skrive disse ut avrundet til én desimal og med passende ledetekster.

Som en test for å sjekke at programmet virker som det skal, kan du bruke data som er gitt i eksempel 6.10 på side 250 om antall tilfeller av salmonella i Norge i perioden 2002–2010. Disse er:

1495, 1537, 1545, 1488, 1805, 1649, 1942, 1234, 1367

Korrekte svar for disse er:

Punktestimat for standardavvik: 214.3

90 % konfidensintervall: [153.9, 366.7]

95 % konfidensintervall: [144.8, 410.6]

99 % konfidensintervall: [129.4, 522.8]

Bruk til slutt programmet til å beregne et 92 % konfidensintervall for standardavviket til antall salmonellatilfeller, og legg ved resultatet du finner.