Øving 2 i ST2202 Anvendt statistikk for vekene 38-39.

Dette er ei øving om resampling ved bootstrap.

Oppgåve 1.

Trekk ut 10 verdiar frå ei normalfordeling. Estimer forventing, standardavvik og varians med tilhøyrande standardavvik v. h. a. bootstrapping. Finn og 95% persentilintervall og Bca-intervall for parameterane. Samanlikn estimata for forventing, standardavvik og varians med dei du får frå det opprinnelege utvalet. Finn og eit konfidensintervall for forventing frå dette utvalet ved å bruke vanleg teori.

Hjelp. Det er lettaste å gjere dette i Splus eller R framfor MINITAB.

Eit eksempel på kommandosekvens som utfører dette er gitt nedanfor.

```
>  x <- rnorm(10, 2, 1)     trekkjer ut 10 verdiar frå ei normalfordeling med forv 2 og standardavvik 1.
> bob1 <- bootstrap(x, stdev)
> summary(bob1)
> plot(bob1)
```

Skal du finne estimat for forventing byter du *stdev* med *mean*. Tilsvarande byter du ut *stdev* med *var* når du vil estimere variansen.
Dersom du ikkje spesifiserer det, vil Splus automatisk gjere 1000 bootstraputval. Du kan spesifisere kor mange du vil gjere ved å bruke B = inne i bootstrap kommandoen. Såleis vil du ved kommandoen nedanfor få gjort 2000 boootstraputval som du kan bruke til å finne estimat for variansen med tilhøyrande standardavvik og konfidensintervall.

```
bob1 <- bootstrap(x, var, B=2000).
```

Oppgåve 2.

Trekk ut 10 verdiar frå eksponentialfordeling med forventing 2. (x <- rexp(10, 0.5)). Estimer forventing, standardavvik og varians med tilhøyrande standardavvik v. h. a. bootstrapping. Finn og 95% persentilintervall og BCa - intervall for parameterane. Samanlikn estimata for forventing, standardavvik og varians med dei du får frå det opprinnelege utvalet. Finn og eit 95% konfidensintervall for forventinga frå dette utvalet ved å bruke vanleg normalfordelingsteori. Finn og eit eksakt konfidensintervall for forventinga og samanlikn konfidensintervallestimata med dei 3 metodane.

Hint: For eksponentialfordelte variable $T_1, T_2, \cdots, T_n$ kan eit finne teoretisk korrekt konfidensintervall ved å bruke at $2\lambda T_i$ er $\chi^2(2)$ (kjikvadratfordelt med 2 fridomsgrader). Her er $E(T_i) = \frac{1}{\lambda}$. Det betyr at $2\lambda\sum_{i=1}^{n} T_i$ er $\chi^2(2n)$. Sannsynsmaksimeringsestimatoren for forventinga er gitt ved $\overline{T}$ og $2n\lambda\overline{T}$ er åltså $\chi^2(2n)$.

Oppgåve 3.

La $X \sim N(2,1)$ og $Z \sim N(0,1)$. La vidare $Y = 5 + 2X + Z$. Finn fordelinga til Y.
Finn korrelasjonen mellom X og Y. Kva blir fordelinga til $Y\big|_{X=x}$?

Trekk 10 verdiar frå fordelinga til X og 10 verdiar frå fordelinga til Z og bruk desse til å generere 10 verdiar frå fordelinga til Y. Estimer korrelasjonen mellom X og Y (korrelasjonsjoeffisienten) og bruk bootstrap til å finne eit estimat for standardavviket til estimatoren for korrelasjonskoeffisienten. Finn og eit 95% konfidensintervall for korrelasjonskoeffisienten.

Kommandoar for bootstrapping.

```
> bdat <- data.frame(x,y)
> bob2 <- bootstrap(bdat, cor(x,y))
```

Resten er som i oppgåve 1.

Oppgåve 4.

Med dei same dataane som i oppgåve 3 føreta enkel lineær regresjon med Y som respons og x som forklaringsvariabel. Kommandoar:

```
> xlm <- lm(y~x)
> summary(xlm)
```

Føreta så ei bootstrap estimering av regresjonskoeffisienten og samanlikn med utskrifta ovanfor. Kommandoar:

```
> bob3 <- bootstrap(bdat, coef(lm(y ~ x)))
```

Resten som i oppgåve 1.