ECON2130 - Utsett eksamen

Vår 2025

Oppgåve 1 - Skatteeksperiment (45%)

Vi ønskjer å gjennomføre eit eksperiment for å undersøke effekten av lågare skatt på arbeidstilbodet til unge i Noreg. Vi trekk ut 1000 tilfeldige personar mellom 20 og 35 år der halvparten vert plassert i ei behandlingsgruppe med lågare skatt og den andre halvparten vert plassert i ei kontrollgruppe.

Inntektene til personane i behandlingsgruppa etter eksperimentperioden er 500 uavhengige stokastiske variablar, $Y_1^B, Y_2^B, ..., Y_{500}^B$, kor $E(Y_i^B) = \mu_B$ og $Var(Y_i^B) = \sigma_B^2$ for alle i.

På same måte føreset vi at inntektene i kontrollgruppa, $Y_1^K, Y_2^K, ..., Y_{500}^K$, er uavhengige stokastiske variablar kor $E(Y_i^K) = \mu_K$ og $Var(Y_i^K) = \sigma_K^2$ for alle i.

Effekten vi ønskjer å estimere er $\beta = \mu_B - \mu_K$. For å estimere β brukar vi estimatoren $\hat{\beta} = \bar{Y}^B - \bar{Y}^K$

a) Kva er forventninga og variansen til $\hat{\beta}$? Er estimatoren forventningsrett?

$$E(\hat{\beta}) = E(\bar{Y}^B - \bar{Y}^K) = E(\frac{1}{N}\sum_i Y_i^B) - E(\frac{1}{N}\sum_i Y_i^K) = \frac{1}{N}N\mu_B - \frac{1}{N}N\mu_K = \mu_B - \mu_K = \beta$$

Estimatoren er forventningsrett.

$$Var(\hat{\beta}) = Var(\bar{Y}^B - \bar{Y}^K) = Var(\frac{1}{N}\sum_i Y_i^B) + Var(\frac{1}{N}\sum_i Y_i^K) = \frac{1}{N^2}N\sigma_B^2 + \frac{1}{N^2}N\sigma_K^2 = \frac{\sigma_B^2 + \sigma_K^2}{N} = \frac{\sigma_B^2 + \sigma_K^2}{500}$$

b) Forklar kort kvifor du kan føresette at $\hat{\beta}$ er tilnærma normalfordelt.

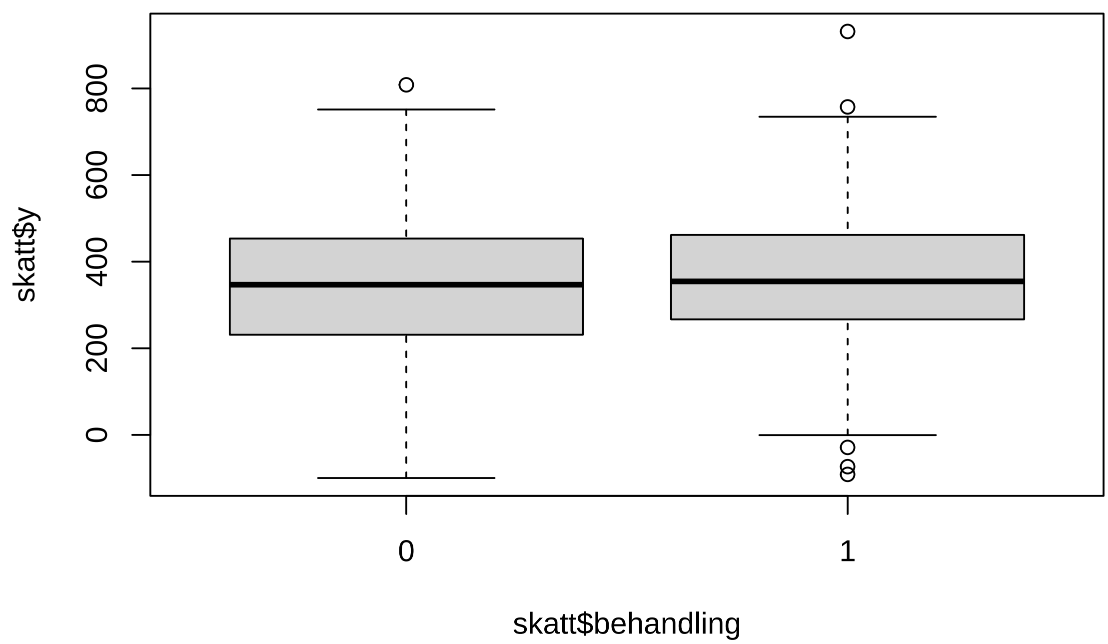
$\bar{Y}^B$ og $\bar{Y}^K$ er gjennomsnitt av mange uavhengige stokastiske variablar. Vi kan då bruke sentralgrenseteoremet som tilseier at desse er tilnærma normalfordelte. Sidan summen av normalfordelte variablar er normalfordelte er $\bar{Y}^B - \bar{Y}^K$ tilnærma normalfordelt.

I fila *skatt_simulert.csv* er det data frå eit slikt tenkt eksperiment. Kvar rad i datasettet er ein observasjon. Datasettet har to variablar:

- y - inntekta for kvar observasjon etter eksperimentet er over
- *behandling* - behandlingsstatus. Den er lik 1 for observasjonane i behandlingsgruppa og 0 for observasjonane i kontrollgruppa.

c) Last inn datasettet og skriv eit R-script som lagar eit boksplott med y på y-aksen og *behandling* på x-aksen. Forklar kva eit boksplott er og kva vi kan lese ut av programmet.

```
skatt <- read.csv('skatt_simulert.csv')
boxplot(skatt$y~skatt$behandling)
```



Den tjukkaste linja er medianen. Verdiane innanfor boksen er mellom første og tredje kvartil. Plottet viser at inntektene i behandlingsgruppa har noko høgare median og verdiane er meir sentrert rundt medianen.

d) Finn eit estimat for β og $SE(\hat{\beta})$.

Estimatet for β:

```
mean(skatt$y[skatt$behandling==1])-mean(skatt$y[skatt$behandling==0])
```

```
## [1] 20.28767
```

Standardfeilen er kvadratrota til variansen. Vi set inn for dei observerte verdiane av σ_B^2 og σ_K^2 i uttrykket frå a):

```r
sqrt(var(skatt$y[skatt$behandling==1])+var(skatt$y[skatt$behandling==0]))/sqrt(500)
```

```
## [1] 9.817592
```

e) Lag eit 85% konfidensintervall for β. Forklar kort korleis ein skal tolke konfidensintervallet.

Vi brukar R til å konstruere konfidensintervallet:

```r
t.test(skatt$y[skatt$behandling==1],
       skatt$y[skatt$behandling==0],
       conf.level = 0.85)
```

```
##
##  Welch Two Sample t-test
##
## data:  skatt$y[skatt$behandling == 1] and skatt$y[skatt$behandling == 0]
## t = 2.0665, df = 996.09, p-value = 0.03904
## alternative hypothesis: true difference in means is not equal to 0
## 85 percent confidence interval:
##   6.144031 34.431308
## sample estimates:
## mean of x mean of y
##  363.3966  343.1089
```

Viss vi gjennomfører mange slike eksperiment og konstruerer eit 85% konfidensintervall kvar gong, vil intervallet i 85% av tilfella innehalde den sanne verdien til β.

Frå oppgåve e)-g) gir det like mykje uttelling å svare med R som for hand.

f) Gjennomfør ein einsidig hypotesetest der du testar nullhypotesa "β er mindre eller lik 0" med 1% signifikansnivå.

$$H_0 : \beta \leq 0, \quad vs \quad H_1 : \beta > 0$$

Vi gjennomfører testen i R:

```r
t.test(skatt$y[skatt$behandling==1],
       skatt$y[skatt$behandling==0],
       alternative = 'greater')
```

```
##
##  Welch Two Sample t-test
##
## data:  skatt$y[skatt$behandling == 1] and skatt$y[skatt$behandling == 0]
## t = 2.0665, df = 996.09, p-value = 0.01952
## alternative hypothesis: true difference in means is greater than 0
## 95 percent confidence interval:
##  4.124135      Inf
## sample estimates:
## mean of x mean of y
##  363.3966  343.1089
```

Sidan p-verdien er større enn 0.01 forkastar vi ikkje H_0.

g) Finn p-verdien for testen i f). Gje ei kort tolking av p-verdien. Er p-verdien til ein hypotesetest ein stokastisk variabel?

Vi fann i f) at p-verdien var lik 0.02. P-verdien er det lågaste signifikansnivået vi forkastar H_0 ved.

Sidan utvalet vårt er stokastisk vil utfallet av hypotesetesten vere stokastisk og dermed også p-verdien. P-verdien er ein stokastisk variabel.

Oppgåve 2 - Sannsyn (20%)

A og B er to av fleire utfall av eit eksperiment. A^c er utfallet "A skjer ikkje". Du får oppgitt følgjande sannsyn:

- $P(A \cup B) = 0.7$
- $P(A \cap B) = 0.1$
- $P(A^c \cap B) = 0.2$

a) Finn $P(B)$

$$P(B) = P(A \cap B) + P(A^c \cap B) = 0.1 + 0.2 = 0.3$$

b) Finn $P(A)$.

$$P(A \cup B) = P(A) + P(B) - P(A \cap B) \iff P(A) = P(A \cup B) + P(A \cap B) - P(B)$$

$$P(A) = 0.7 + 0.1 - 0.3 = 0.5$$

c) Er A og B uavhengige hendingar?

$$P(A) * P(B) = 0.15 \neq P(A \cap B)$$

A og B er ikkje uavhengige.

d) Finn $P(A|B)$ og $P(B|A)$.

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{0.1}{0.3} = 0.33$$

$$P(B|A) = \frac{P(A \cap B)}{P(A)} = \frac{0.1}{0.5} = 0.2$$

Oppgåve 3 - Mål (35%)

Vi kallar talet "mål i ein tilfeldig vald eliteseriekamp" i fotball for X. Vi føreset at X er Poisson-fordelt med $\lambda = 2.7$.

a) Kva inneber det å føresette at X er Poisson-fordelt? Diskuter kort om det er rimeleg å føresette Poisson-fordeling i dette tilfellet.

X er Poisson-fordelt viss:

- Antal mål per kamp er uavhengig av antal mål i andre kampar.
- Forventa antal mål per kamp er lik λ
- To mål kan ikkje kome nøyaktig samtidig

Av desse føresetnadane er det dei to første som er mest tvilsame i dette tilfellet. Eit mål i ein kamp kan påverke sannsynet for mål i andre kampar og det er heller ikkje gitt at λ er konstant innad eller mellom kampar.

b) Finn $P(X \geq 7)$.

$$P(X \geq 7) = 1 - P(X \leq 6)$$

Vi finn svaret i R:

```
1-ppois(6,2.7)
```

```
## [1] 0.02056945
```

c) I løpet av ein sesong er det til saman 240 kampar. Du kallar antal mål i løpet av ein sesong for Y. Kva er fordelinga til Y?

Summen av n Poisson-fordelte variablar er Poisson-fordelt med $\lambda = \lambda_1 + \lambda_2 + ... + \lambda_n$

Derfor er $Y \sim Pois(240 * 2.7)$.

d) Finn $P(Y < 600)$

$$P(Y < 600) = P(Y \leq 599)$$

Vi finn svaret i R:

```r
ppois(599,2.7*240)
```

```
## [1] 0.02722954
```

Her er det også innanfor å bruke ei normaltilnærming.

e) Kva er sannsynlegheita for at gjennomsnittleg antal mål per kamp i ein sesong er større enn 3?

$$P(Y/240 > 3) = P(Y > 720) = 1 - P(Y \leq 719)$$

Vi finn svaret i R:

```r
1-ppois(719,2.7*240)
```

```
## [1] 0.002839127
```