```r
knitr::opts_chunk$set(echo = FALSE)
```

Oppgave 1

Vi har den lineære regressjonen $y_i = \beta_0 + \beta_1 x_i + \epsilon_i$

y_i er den avhengige variabelen.
x_i er den uavhengige variabelen.
β_0 er konstant leddet.
β_1 er koeffisienten.
ϵ_u er feilleddet/støyleddet, vi antar for en tilfeldig i så er$E(\epsilon_i) = 0$

Vi ønsker å finne y_i kurven som minimerer summen av den kvadrerte residualene (Ordinary Least squares). Vi kaller den faktisk sanne verdien av den avhengige vairabelen for y og resultat fra våres etsimator for $\hat{y}$

Vi ønsker å minimere $\sum_{i=1}^{n}(y_i - \hat{y_i})^2 = OLS$

Vi har to paramtere som vi må bestemme i $\hat{y}$, nemlig β_0 og β_1.

For å minimere disse må vi ta den partiellderiverte av både β_0 og β_1 og sette dem lik 0.

$\frac{\partial \text{OLS}}{\partial \beta_0} = -2\sum_{i=1}^{n}(y_i - \beta_0 - \beta_1 x_i) = 0$ $\frac{\partial \text{OLS}}{\partial \beta_1} = -2\sum_{i=1}^{n}(y_i - \beta_0 - \beta_1 x_i)x_i = 0$

For å vite at parameteren vi får, $\hat{\beta}_0$ og $\hat{\beta}_1$, faktisk minimerer de kvadrerte residualene sjekker vi andreordensbetingelsen.

$$Hessian(OLS) = \begin{bmatrix} \frac{\partial^2 \text{OLS}}{\partial \beta_0^2} & \frac{\partial^2 \text{OLS}}{\partial \beta_0 \partial \beta_1} \\ \frac{\partial^2 \text{OLS}}{\partial \beta_1 \partial \beta_0} & \frac{\partial^2 \text{OLS}}{\partial \beta_1^2} \end{bmatrix} \geq 0$$

Dette forsikrer oss om at Hessematrisen er positive semidefinite hvilket betyr at OLS funksjonen er konveks og dermed at vi minimerer de kvadrerte residualene.

Vi er interessert i å finne $\hat{\beta}_0$ og $\hat{\beta}_1$. Vi har fra før:

$\frac{\partial \text{OLS}}{\partial \beta_0} = -2\sum_{i=1}^{n}(y_i - \beta_0 - \beta_1 x_i) = 0$
$\frac{\partial \text{OLS}}{\partial \beta_1} = -2\sum_{i=1}^{n}(y_i - \beta_0 - \beta_1 x_i)x_i = 0$

hvilket vi kan skrive som:

$(1)\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{X}$
$(2)\hat{\beta}_1 = \frac{\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})}{\sum_{i=1}^{n} X_i - \bar{X}}$

Her er (1) den første momentbetingelsen og (2) den andre momentbetingelsen bare løst for β_0 og β_1 respektivt.

Vi har tre antakelser:

1)$E(u|x) = 0$
2) i.i.d, altså at utvalget er uavhengig og identiske distrbuerte til den faktiske populasjonen.
3)$0 < E(x^4) < \infty$ og $0 < E(y^4) < \infty$, X og Y har begrenset kurtosis. Med andre ord kan vi si at store ekstremverdier er usannsynlig

Hvis disse holder så finner vi:

$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} x_i$
$\bar{Y} = \frac{1}{n}\sum_{i=1}^{n} y_i$

Når vi har $\bar{Y}$ og $\bar{X}$ kan vi enkelt finne $\hat{\beta_1}$ fra andre momentbetingelse (2) og deretter $\hat{\beta_0}$ fra første momentbetingelse (1).

Slik finner vi $\hat{Y} = \hat{B_0} + \hat{B_1}X$

Oppgave 2

a)

En dummyvariabel, også kalt indikatorvariabel, er en binær variabel som tar verdien 0 eller 1 for å indikere tilstedeværelsen av en effekt/variabel eller ikke tilstedeværelse. En dummyvariabel kan være student med 0 for nei og 1 for ja når vi undersøker inntektselastisitet i en befolkningen.

Når den uavhengige variabelen er kontinuerlig representerer koeffisienten endringen i den uavhengige variabelen på den avhengige variabelen. Det forteller oss noe om sammenhengen mellom den avhengige og uavhengige variabelen. Et eksempel på dette kan sammenhengen mellom utdanningsnivå og inntekt. Inntekt er en avhengige variabel som avhenger av den uavhengige variabelen utdanningsnivå og koeffisenten representerer endringen av et ekstra år med utdanning på inntekt.

Når den uavhengige variabelen er binær representerer koeffisenten effekten av tilstedeværelsen av den uavhengige variabel på den avhengige variabelen. Det sier noe om effekten og forksjellen mellom tilstedeværelsen av den uavhengige variabelen på den avhengige variabelen. Et eksempel på dette kan være sammenhengen mellom inntekt og hvorvidt et individ har høyere utdanning eller ikke. Inntekt er en avhengige variabel som avhenger av den uavhengige og binære variabelen for høyere utdanning. Hvis ingen høyere utdanning så er den uavhengige variabelen 0 og vice versa. Koeffisienten representerer effekten av høyere utdanning på inntekt.

Dersom den uavhengige variabelen er binær må vi lage konfidensintervall på en annen måte. For de kontinuerlige uavhengige variabelen er et konfidensintervall et mål på usikkerheten rundt verdien på koeffisenten gitt en p-verdi (Eks, p=0.05 for 95% konfidensintervall).For den binære uavhengige variabelen måler vi differeansen mellom gjennomsnittsverdien med tilstedeværelse av den uavhengige variabelen og gjennomsnittsverdien uten tilstedeværelse. Konfidensintervallet måler så usikkerheten rundt forskjellen mellom de to gjennomsnittsverdiene gitt en p-verdi.

Under hypotesetesting kan nullhypotesen H_0 og alternativhypotesen $H_a$se ganske lik ut for både en regressjon med binær - og kontinuerlig uavhengig variabel. I begge regresjonen bruker vi β som koeffisient. Vi sier at:
$H_0 : \beta = 0$
$H_a : \beta \neq 0$
Her blir tolkningen annerledes. Hypotesting for en regressjon med kontinuerlig uavhengig variabel tester om en liten endring i den uavhengige variabelen har en signifikant effekt på den avhengige variabelen. Hypotesting for en regressjon med binær uavhengig variabel tester om en det er en signifikant forskjell mellom tilstedeværelse eller ikke av den uavhenige variabel på den avhengige variabelen.

b)

Vi har fra oppgaven: $\ln(Y_i) = \beta_0 + \beta_1 D + u_i$

Vi skal tolke β_1 i denne modellen både tilnærmet og eksakt.

Når β_1 er "liten", for eksempel $\beta_1 \leq 0.2$ kan vi tolke β_1 som prosentvis endringen når vi har tatt logaritem av Y_i med tilstrekkelig nøyaktighet. Altså en $\beta_1 = 0.1$ vil da være en 10% økning i lønn dersom en er kvinne $(D = 1)$

Derimot for en eksakt tolkning av $\beta = 0.1$ må vi eksponesiere begge sider.

$Y_i = e^{0.1} \approx 1.1052$ For å finne faktisk prosentvis økning må vi trekke fra den gamlle verdien, dele på den gamle verdien, og gange med 100

$\%\Delta Y_1 = \frac{(e^{0.1} - 1)}{1} * 100 = 0.1052$

Vi ser at den eksakte og tilnærmete tolkningen er veldig like for små nok verdier av β_1

c)

Vi har $Q = AL^\alpha K^\beta$

Vi transfomerer funksjonen ved å ta den naturlige logaritmen av den; $\ln(Q) = \ln(A) + \alpha \ln(L) + \beta \ln(K)$.

Vi har nå en lineær modell for parametrene α og β. Vi kan bruke lineær regresjon for å estimere α og β og konstanten $\ln(A)$ som vi kan kalle C.
$\ln(Q) = \alpha \ln(L) + \beta \ln(K) + C$

Vi kan med OLS metoden beskrevet i oppgave 1 finne de nøvendige estimatene:
$\hat{\alpha}, \hat{\beta}, \hat{C}$,slik at $ln(Q) = \hat{\alpha} \ln(L) + \hat{\beta} \ln(K) + \hat{C}$

I dette tilfellet blir $\hat{\alpha}$ den estimerte produksjonselastisiteten for arbeid L og $\hat{\beta}$ den estimerte produksjonselastisiteten for kapital K og $\hat{C}$ konstantleddet.

Anta at estimatene vi fikk ble $ln(Q) = 0.5 * ln(L) + 0.4 * ln(K)$ så vil en 1% økning i arbeid L gi 0.5% i produksjon Q og 1% økning i kapital K gi 0.4% økning i produksjon Q.

Ved flere typer forskjellige maskiner og typer kapital ville vi utvidet kapital $K = \sum_i^n k_i$ og arbeid $L = \sum_i^n l_i$ hvor det er n typer kapital og arbeid. På samme måte får vi også $\alpha = \sum_i^n \alpha_i$ og $\beta = \sum_i^n \beta_i$ slik beholder vi de samme egenskapene som originale modellen bare at nå kan vi isolere effekten av en spesifikt type kapital eller arbeid på produksjon Q.

Oppgave 3

a)

Vi har $\widehat{\text{Lønn}} = 10.73 + 1.78 * \text{Mann}$ hvor

$$\text{Mann} = \begin{cases} 1, & \text{hvis mann} \\ 0, & \text{ellers} \end{cases}$$