

Er det høyde som bestemmer inntekt?

Assignment 2 - MSB 105

Ole Alexander Bakkevik & Sindre M. Espedal

Introduksjon

I dette assignmentet skal vi besvare på problemstillingen “*er det høyde som bestemmer inntekt?*”. Vi benytter datasettet **heights** i pakken **modelr** for å svare på problemstillingen.

Litteraturgjennomgang

Det argumenteres for at fysiske egenskaper spiller en sentral rolle i interaksjoner og resultater i arbeidslivet og det er aktiv litteratur som fokuserer på hvordan attraktivitet, vekt og kroppsbilde påvirker interaksjoner og resultater på arbeidsplassen (Judge og Cable, 2004). For eksempel høyere individer vurderes som mer overbevisende (Young og French, 1996) og mer attraktive som kamerater (Freedman, 1980).

Hensley noterte seg at oppfatningen av at høyere individer er på en eller annen måte mer kapable, dyktige eller mer kompetente ser ut til å stemme (Hensley, 1993). Denne påstanden kan styrkes i det Lester og Sheehan fant ut at sjefene sin forventning var at korte politifolk skulle motta flere klager, forårsake flere disiplinære problemer og skape dårligere moral enn det høyere politifolk ville gjort (Lester og Sheehan, 1980).

Studier har vist at folk oppfatter verdifulle ting som større enn mindre verdifulle ting; for eksempel oppfattes mynter som er større enn pappskiver med identisk diameter (Judge og Cable, 2004). Denne skjevheten strekker seg også til vurderingen om individers høyde og verdsettelse.

Dette styrkes gjennom en studie av kanadiske velgere, der de viste at etter valget i 1988, bedømte velgerne vinneren (Brian Mulroney) til å være høyere enn før valget. I tillegg bedømte velgerne taperne til å være kortere enn før valget (Judge og Cable, 2004). Høyde er også en metafor for betydningen av makt (Judge og Cable, 2004) og er ofte brukt som en “*heuristikk for dominans*” (Young og French, 1998). I språket er også høyde av sosial verdi. Når en person er høyt aktet, kan han beskrives som en “*stor mann*”, og vi “*Ser opp*” til og beundrer de høye individene (Frieze et al., 1990).

I studien “*The Effect of Physical Height on Workplace Success and Income: Preliminary Test of a Theoretical Model*” fokuserer de på hvilken rolle egenskapen høyde har på suksess i arbeidslivet. I studien kommer det frem at resultatene som er presentert i artikkelen tyder tydelig på at fysisk høyde påvirker folk karrier og interaksjoner på arbeidsplassen og er derfor verdig til fortsatt vitenskapelig undersøkelse (Judge og Cable, 2004).

Datasettet

Som nevnt bruker vi datasettet **heights** i pakken **modelr**, men siden dataene er amerikanske, er måleenhetene av amerikanske verdier. Vi konverterer dermed først til det metriske systemet som vi bruker i Norge slik at resultatene av analysene skal gi bedre forståelse. I tillegg konverterer vi inntekten fra dollar til norske kroner.

```
#skifter navn på datasettet, fra heights til hoyde
hoyde <- heights
```

```
#konverterer måleenhetene og valutaen
hoyde$hoyde_cm <- hoyde$height*2.54
hoyde$vekt_kg <- hoyde$weight*0.45359237
hoyde$inntekt_kr <- hoyde$income*8.37 #Gjennomsnittlig valutakurs i uke 42, 2021.
hoyde$bmi <- I(hoyde$vekt_kg/(hoyde$hoyde_cm)^2*(10000))
```

I tillegg forenkler vi variabelen *marital*. Vi forenkler dette ved å kun benytte gift, eller ikke-gift istedenfor flere alternativer som f.eks *singel*, *skilt*, *enke* osv.

```
hoyde <- hoyde %>%
mutate(
  married = factor(
    case_when(
      marital == 'married' ~ TRUE,
      TRUE ~ FALSE
    )
  )
)
```

Nå er sivistatusen *married* satt som TRUE, og alle andre kategoriene i den nye variabelen er FALSE.

Beskrivende statistikk

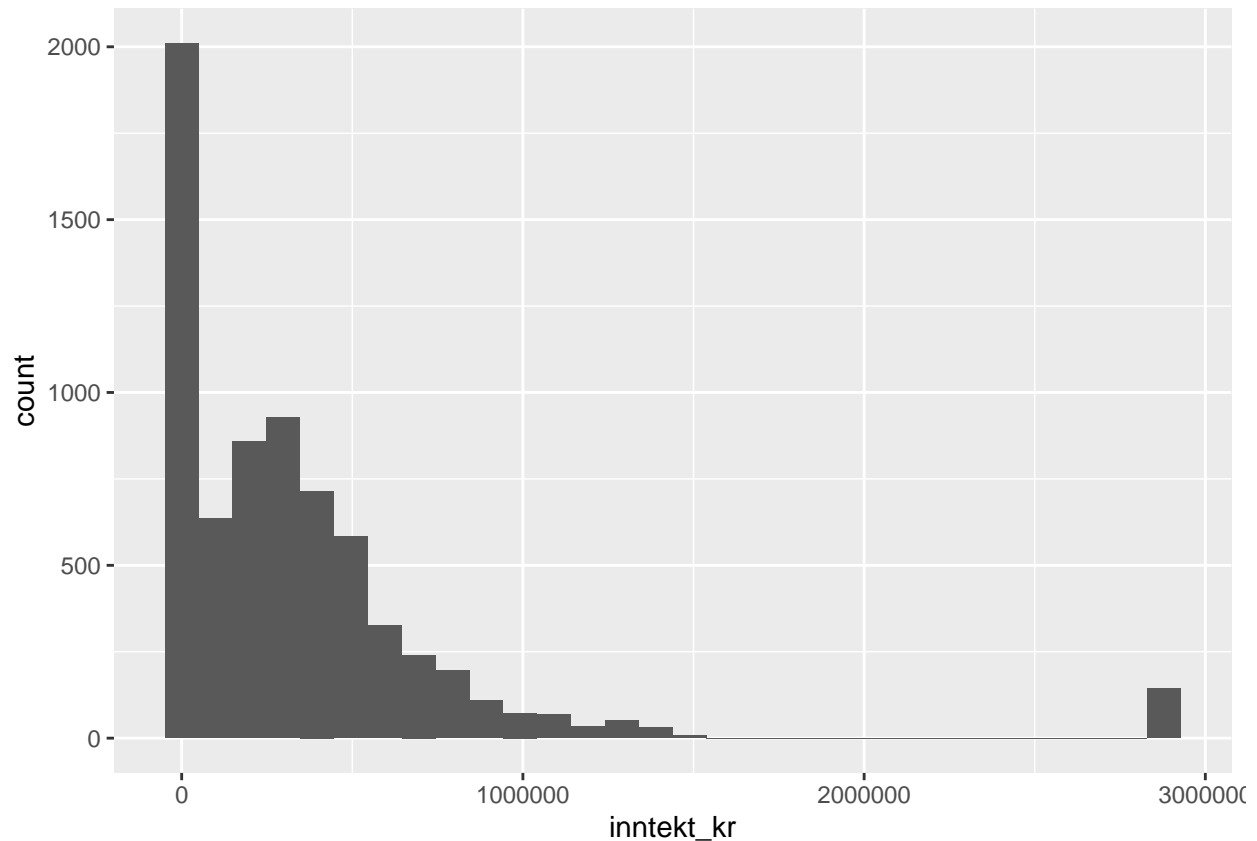
Denne delen vet jeg vertfall at Ole Alexander har gjort.

EDA

Tror Ole Alexander har gjort dette også. Gjør den kort og kjapt uansett.

```
#Histogram av variabelen inntekt_kr
ggplot(aes(x = inntekt_kr), data = hoyde) + geom_histogram()
```

```
## 'stat_bin()' using 'bins = 30'. Pick better value with 'binwidth'.
```



Uteliggerne til høyre skyldes at det er noen få som tjener veldig mye. Disse er summert sammen og utregnet gjennomsnittet på dem. Dette er på grunn av personverns-årsaker.

```
#personer uten inntekt
sum(hoyde$inntekt_kr == 0)
```

```
## [1] 1740
```

Vi har 1 740 observasjoner som har inntekt lik 0.

Regresjonsanalyse

Alle nye- og forenklede variabler er lagt inn fra før. Det samme gjelder konverteringene.

Den første modellen er en veldig enkel regresjonsmodell som ser kun på hvordan høyden påvirker inntekten.

```
modell_1 = "inntekt_kr ~ hoyde_cm"
lm_1 <- lm(modell_1, data = hoyde)
summary(lm_1)
```

```
##
## Call:
## lm(formula = modell_1, data = hoyde)
##
```

```
## Residuals:
##      Min       1Q   Median       3Q      Max
## -766554 -263745  -91173   124563  2685330
##
## Coefficients:
##              Estimate Std. Error t value      Pr(>|t|)
## (Intercept) -1329893.0    89841.5  -14.80 <0.0000000000000002 ***
## hoyde_cm      9825.9      526.1    18.68 <0.0000000000000002 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 456600 on 7004 degrees of freedom
## Multiple R-squared:  0.04744,    Adjusted R-squared:  0.0473
## F-statistic: 348.8 on 1 and 7004 DF,  p-value: < 0.00000000000000022
```

Ut i fra denne enkle regresjonen i **modell_1** ser vi at hvis vi øker høyden med én ekstra centimeter, så øker årlig lønn med 9825.9kr. Den uttrykkes også som signifikant, men forklaringsvariansen er kun på 4,7% som vil si at 95,3% av modellen kan ikke forklares gjennom regresjonen (*u*). Derfor er dette egentlig en veldig dårlig modell.

I neste modell legger vi til en ekstra variabel, variabelen *vekt_kg*, for å se hvordan dette påvirker regresjonsmodellen.

```
modell_2 = "inntekt_kr ~ hoyde_cm + vekt_kg"
lm_2 <- lm(modell_2, data = hoyde)
summary(lm_2)
```

```
##
## Call:
## lm(formula = modell_2, data = hoyde)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -828525 -258745  -90442   124446  2701163
##
## Coefficients:
##              Estimate Std. Error t value      Pr(>|t|)
## (Intercept) -1441826.3    93195.8  -15.471 < 0.0000000000000002 ***
## hoyde_cm      11226.7      600.5    18.696 < 0.0000000000000002 ***
## vekt_kg       -1481.3      309.0    -4.794    0.00000167 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 458000 on 6908 degrees of freedom
## (95 observations deleted due to missingness)
## Multiple R-squared:  0.05077,    Adjusted R-squared:  0.0505
## F-statistic: 184.7 on 2 and 6908 DF,  p-value: < 0.00000000000000022
```

I **modell_2** påvirker *vekt_kg* negativt på *inntekt_kr*. Hvis *vekt_kg* øker med én ekstra kg, så reduseres årlig *inntekt_kr* 1481.3kr. Denne variabelen er også signifikant, men forklaringsvariansen på overkant av 5% indikerer at **modell_2** er også en dårlig modell, ettersom den har en liten forklaring.

I **modell_3** legges det til en ytterligere variabel, variabelen *bmi*.

```
modell_3 = "inntekt_kr ~ hoyde_cm + vekt_kg + bmi"
lm_3 <- lm(modell_3, data = hoyde)
summary(lm_3)
```

```
##
## Call:
## lm(formula = modell_3, data = hoyde)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -864541 -257416  -91679   124029  2696877
##
## Coefficients:
##              Estimate Std. Error t value    Pr(>|t|)
## (Intercept) -1907489     427914  -4.458 0.0000084174 ***
## hoyde_cm      13972         2534   5.513 0.0000000365 ***
## vekt_kg       -4198         2456  -1.709    0.0874 .
## bmi           7834         7026   1.115    0.2649
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 458000 on 6907 degrees of freedom
## (95 observations deleted due to missingness)
## Multiple R-squared:  0.05094,    Adjusted R-squared:  0.05053
## F-statistic: 123.6 on 3 and 6907 DF,  p-value: < 0.00000000000000022
```

Ut i fra **modell_3** så ser vi at hvis *bmi* øker med én ekstra enhet, så øker årlig *inntekt_kr* med 7834kr.

modell_3 viser 3 variabler hvor *hoyde_cm* er den eneste signifikante variabelen med et konfidensintervall på 95%. Likevel har forklaringsvariansen hatt en minimal økning på en ellers så svært liten forklaringskraft.

Litteraturliste

- Freedman, D. G. (1980). The Social and the Biological: A Necessary Unity. *Zygon*, 15(2), 117–131. <https://doi.org/10.1111/j.1467-9744.1980.tb00381.x>
- Frieze, I. H., Olson, J. E., og Good, D. C. (1990). Perceived and Actual Discrimination in the Salaries of Male and Female Managers. *Journal of Applied Social Psychology*, 20(1), 46–67. <https://doi.org/10.1111/j.1559-1816.1990.tb00377.x>
- Hensley, W. E. (1993). Height as a Measure of Success in Academe. *Psychology: A Journal of Human Behavior*, 30(1), 40–46.
- Judge, T. A., og Cable, D. M. (2004). The Effect of Physical Height on Workplace Success and Income: Preliminary Test of a Theoretical Model. *The Journal of Applied Psychology*, 89(3), 428–441. <https://doi.org/10.1037/0021-9010.89.3.428>
- Lester, D., og Sheehan, D. (1980). Attitudes of Supervisors toward Short Police Officers. *Psychological Reports*, 47(2), 462–462. <https://doi.org/10.2466/pr0.1980.47.2.462>
- Young, T. J., og French, L. A. (1996). Height and Perceived Competence of US Presidents. *Perceptual and Motor Skills*, 82(3 Pt 1), 1002. <https://doi.org/10.1177/003151259608200301>
- Young, T. J., og French, L. A. (1998). Heights of U.S. Presidents: A Trend Analysis for 1948. *Perceptual and Motor Skills*, 87(1), 321–322. <https://doi.org/10.2466/pms.1998.87.1.321>