

# 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 R-pakken *modelr*<sup>1</sup> 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, 2004b). For eksempel hevdes det at høyere individer vurderes som mer overbevisende (Young og French, 1996) og mer attraktive som kamerater (Freedman, 1980).

Hensley (1993) 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. Denne påstanden kan styrkes i det Lester og Sheehan (1980) 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.

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 som større (Judge og Cable, 2004b). 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, 2004b). Høyde er også en metafor for betydningen av makt (Judge og Cable, 2004b) 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).

Judge og Cable (2004a) fokuserer på hvilken rolle egenskapen høyde har for suksess i arbeidslivet. I studien kommer det frem at resultatene tyder på at fysisk høyde påvirker folk karriere og interaksjoner på arbeidsplassen. Man kan derfor hevde at problemstillingen er verdig fortsatt vitenskapelige undersøkelser (Judge og Cable, 2004b).

### Datasettet

Som nevnt bruker vi datasettet **heights** i pakken **modelr**, men siden dataene er amerikanske, er måleenhetene amerikanske enheter. Vi konverterer dermed først til det metriske systemet som vi bruker i Norge slik at resultatene lettere kan forstås. I tillegg konverterer vi inntekten fra dollar til norske kroner.

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

---

<sup>1</sup>(R Core Team, 2021) og (Wickham, 2020)

Nå er sivistatusen *married* satt som TRUE, og alle andre kategoriene i den nye variabelen er FALSE. Da måles det kun om man er gift eller ikke.

```
##      income      height      weight      age
## Min.      : 0.0    Min.    :52.0    Min.    : 76.0    Min.    :47.00
## 1st Qu.: 165.5    1st Qu.:64.0    1st Qu.:157.0    1st Qu.:49.00
## Median : 29589.5   Median :67.0    Median :184.0    Median :51.00
## Mean   : 41203.9   Mean    :67.1    Mean    :188.3    Mean    :51.33
## 3rd Qu.: 55000.0   3rd Qu.:70.0    3rd Qu.:212.0    3rd Qu.:53.00
## Max.    :343830.0   Max.     :84.0    Max.     :524.0    Max.     :56.00
##
##                      NA's      :95
##      marital      sex      education      afqt
## single   :1124    male   :3402    Min.      : 1.00    Min.      : 0.00
## married  :3806    female:3604    1st Qu.:12.00    1st Qu.: 15.12
## separated: 366                      Median :12.00    Median : 36.76
## divorced :1549                      Mean   :13.22    Mean   : 41.21
## widowed  : 161                      3rd Qu.:15.00    3rd Qu.: 65.24
##
##                      Max.      :20.00    Max.      :100.00
##                      NA's      :10      NA's      :262
```

	income	height	weight	age	marital	sex	education	af
	Min. : 0.0	Min. :52.0	Min. : 76.0	Min. :47.00	single :1124	male :3402	Min. : 1.00	M
	1st Qu.: 165.5	1st Qu.:64.0	1st Qu.:157.0	1st Qu.:49.00	married :3806	female:3604	1st Qu.:12.00	1s
	Median : 29589.5	Median :67.0	Median :184.0	Median :51.00	separated: 366	NA	Median :12.00	M
	Mean : 41203.9	Mean :67.1	Mean :188.3	Mean :51.33	divorced :1549	NA	Mean :13.22	M
	3rd Qu.: 55000.0	3rd Qu.:70.0	3rd Qu.:212.0	3rd Qu.:53.00	widowed : 161	NA	3rd Qu.:15.00	3r
	Max. :343830.0	Max. :84.0	Max. :524.0	Max. :56.00	NA	NA	Max. :20.00	M
	NA	NA	NA's :95	NA	NA	NA	NA's :10	N

## Beskrivende statistikk

Variablene i dattasettet **heights**, samt verdiene og målingene er hentet fra helpfunksjonen i R og disse variablene er:

Income

- Årlig inntekt. Topp 2% er gjennomsnittet av dem.

Height

- Høyde, i inches (amerikansk måleenhet)

Weight

- Vekt, i pounds (amerikansk måleenhet)

Age

- Alder, i år, mellom 47 og 56

Marital

- Sivilstatus

Sex

- Kjønn

Education

- År med utdanning

Afqt

- Prosentcore på en “*IQ test*”.

Variablene i datasettet **hoyde** er like, i tillegg til noen ekstra variabler:

inntekt\_kr

- Årlig inntekt, målt i norske kroner.

hoyde\_cm

- Høyde, målt i centimeter.

vekt\_kg

- Vekt, målt i kilogram.

bmi

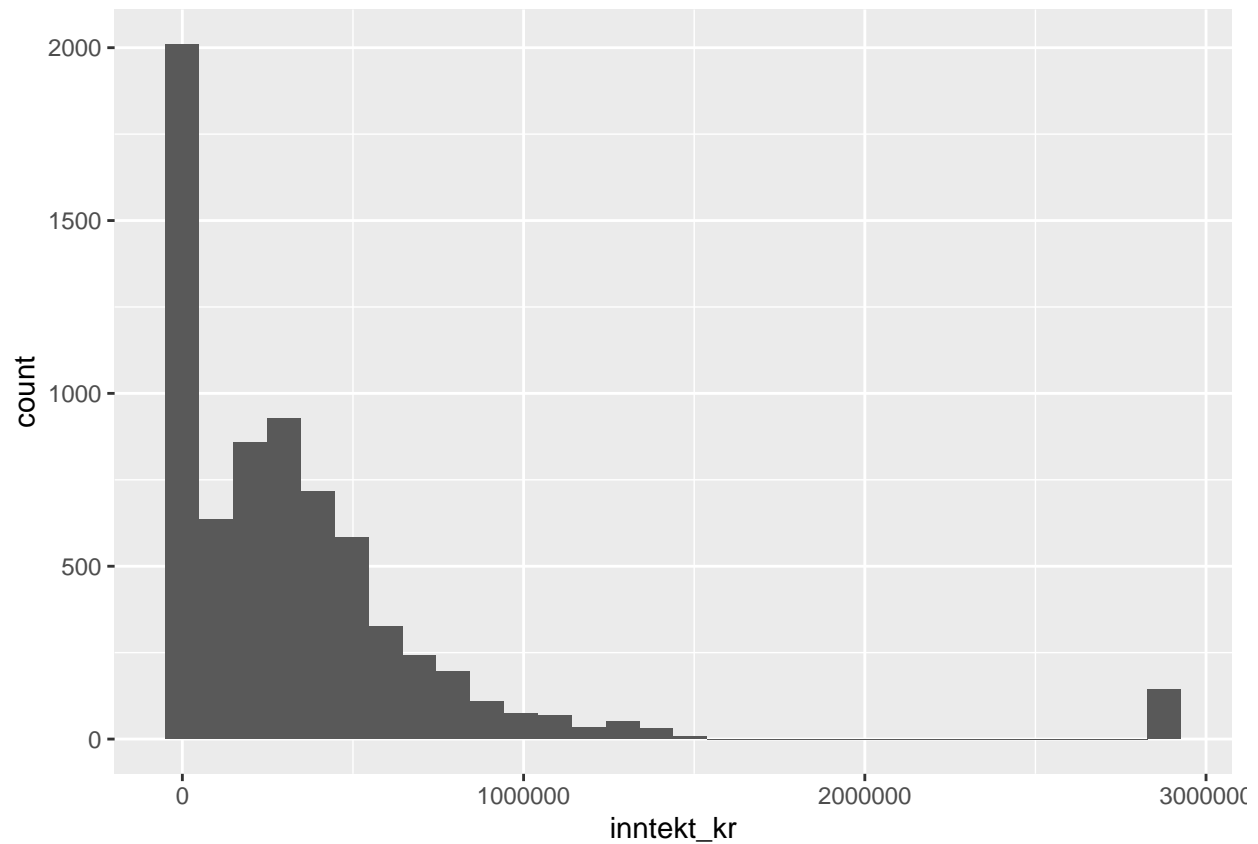
- Body-mass-index

married

- Gift eller ikke-gift

## EDA

Histogram for *inntek\_kr*.



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.

```
## [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.

```
##
## Call:
## lm(formula = modell_1, data = .)
##
## 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  <2e-16 ***
## hoyde_cm      9825.9       526.1   18.68  <2e-16 ***
## ---
## 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: < 2.2e-16
```

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.

```
##
## 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  < 2e-16 ***
## hoyde_cm      11226.7     600.5   18.696  < 2e-16 ***
## 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: < 2.2e-16
```

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. Et annet problem er korrelasjonen mellom høyde og vekt. Høye personer veier mer selv om de er like slanke.

I **modell\_3** legges det til en ytterligere variabel, variabelen *bmi*.

```
##
## 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: < 2.2e-16
```

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.

I **modell\_4** legges variablene *education* og *age* til i regresjonsmodellen.

```
##
## Call:
## lm(formula = modell_4, data = hoyde)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -994470 -230101  -57569   124319  2855576
##
## Coefficients:
##              Estimate Std. Error t value    Pr(>|t|)
## (Intercept) -2603720     413936  -6.290 0.000000000337 ***
## hoyde_cm      13468         2344   5.745 0.000000009591 ***
## vekt_kg       -4736         2272  -2.084    0.0372 *
## bmi           11853         6505   1.822    0.0685 .
## education     68611         1964  34.934 < 2e-16 ***
## age          -3815         2276  -1.676    0.0937 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
```

```
## Residual standard error: 422300 on 6895 degrees of freedom
## (105 observations deleted due to missingness)
## Multiple R-squared: 0.1939, Adjusted R-squared: 0.1933
## F-statistic: 331.7 on 5 and 6895 DF, p-value: < 2.2e-16
```

**modell\_4** viser at *age* påvirker årlig *inntekt\_kr* negativt mens *education* har en positiv effekt. variabelen *education* er også svært signifikant. De 2 ekstra variablene øker forklaringskraften betydelig, helt opp til 19,4%.

```
##
## Call:
## lm(formula = modell_4b, data = hoyde)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1088.83  -246.18    31.39   197.09  1526.94
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   -110.3384   130.1827  -0.848   0.3967
## sexfemale      200.6138   181.2316   1.107   0.2684
## bmi            -1.7248    1.0195  -1.692   0.0907 .
## education      64.7043    2.1332  30.332 < 2e-16 ***
## age           -5.2366    2.3999  -2.182   0.0291 *
## marriedTRUE    217.3119   11.0401  19.684 < 2e-16 ***
## sexfemale:bmi  -0.1717    1.2656  -0.136   0.8921
## sexfemale:education -19.4599   2.9359  -6.628 0.0000000000365 ***
## sexfemale:age    0.2840    3.3472   0.085   0.9324
## sexfemale:marriedTRUE -204.6184   15.3131 -13.362 < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 310.2 on 6891 degrees of freedom
## (105 observations deleted due to missingness)
## Multiple R-squared: 0.2729, Adjusted R-squared: 0.272
## F-statistic: 287.4 on 9 and 6891 DF, p-value: < 2.2e-16
```

Res.Df	Df	F	Pr(>F)
6.9e+03			
6.89e+03	4	4.87	0.000636

Ser at vi kan forkaste  $H_0$  så konklusjonen er at kjønn og alder må være med i modellen.

Forskjellen mellom vanlige og robuste standard errors

```
##
## t test of coefficients:
##
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   -110.33836  143.36350  -0.7696   0.4415
## sexfemale      200.61385  185.00236   1.0844   0.2782
## bmi            -1.72479    1.10700  -1.5581   0.1193
## education      64.70427    2.74230  23.5949 < 2.2e-16 ***
## age           -5.23656    2.63342  -1.9885   0.0468 *
```

```
## marriedTRUE          217.31191    11.75232    18.4910    < 2.2e-16 ***
## sexfemale:bmi         -0.17170     1.29361    -0.1327     0.8944
## sexfemale:education   -19.45987     3.34777    -5.8128    0.000000006418 ***
## sexfemale:age          0.28396     3.39259     0.0837     0.9333
## sexfemale:marriedTRUE -204.61839    15.13167   -13.5225    < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

## Original model not retained as part of coeftest object. For additional model summary information (r.
## This message is displayed once per session.
```

## Huxreg

Setter opp en **Huxtable** for å vise en oversikt over modell 1, 2 og 3.

Variabelen *hoyde\_cm* er signifikant gjennom alle tre modellene mens *vekt\_kg* er kun signifikant i **modell\_2**. Variabelen *hoyde\_cm* får også en større påvirkning på *inntekt\_kr* når flere variabler legges til. Forklaringskraften øker minimalt fra **modell\_1**, men den øker ikke mellom **modell\_2** og **modell\_3**. Forklaringskraften er også veldig liten.

## Interaksjon

Denne modellen har en interaksjon på variabelen *sex*, som vil si en modell for kvinner og menn.

```
##
## Call:
## lm(formula = modell_int, data = hoyde)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -844832 -243051  -90336   125765  2664145
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   -2296981.00   3061743.05  -0.750    0.453
## sexfemale      364907.00   3877913.34   0.094    0.925
## hoyde_cm       14177.21    17010.78   0.833    0.405
## vekt_kg         217.78     29577.95   0.007    0.994
## I(vekt_kg^2)    -29.54       71.13  -0.415    0.678
## bmi            17853.31    96322.70   0.185    0.853
## I(bmi^2)        -64.14       767.00  -0.084    0.933
## sexfemale:hoyde_cm   -31.35    22460.23  -0.001    0.999
## sexfemale:vekt_kg   -18273.48   40602.63  -0.450    0.653
## sexfemale:I(vekt_kg^2)  67.20     105.64   0.636    0.525
## sexfemale:bmi       25515.27   120895.62   0.211    0.833
## sexfemale:I(bmi^2)   -193.30     937.66  -0.206    0.837
##
## Residual standard error: 455700 on 6899 degrees of freedom
## (95 observations deleted due to missingness)
## Multiple R-squared:  0.06166,    Adjusted R-squared:  0.06016
## F-statistic: 41.21 on 11 and 6899 DF,  p-value: < 2.2e-16
```

Vi ser fra regresjonsmodellen at dummyen for *sexfemale* og interaksjonsvariablene ikke er signifikante.

# Hypotesetesting

## Test av koeffisientene samt korrelasjon

Vi kjører test av koeffisientene i interaksjonsmodellen mellom kvinner og menn.

```
##
## t test of coefficients:
##
##           Estimate Std. Error t value      Pr(>|t|)
## (Intercept) -2603720.2   449564.5 -5.7917 0.000000007276 ***
## hoyde_cm      13468.1     2534.7  5.3136 0.000000110870 ***
## vekt_kg       -4736.3     2379.1 -1.9908    0.04654 *
## bmi           11852.6     6433.8  1.8422    0.06548 .
## education     68611.2     2678.0 25.6207 < 2.2e-16 ***
## age           -3814.6     2315.2 -1.6477    0.09947 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

##           2.5 %      97.5 %
## (Intercept) -3415162.4005 -1792278.0593
## hoyde_cm      8872.3928    18063.7622
## vekt_kg       -9190.8220    -281.8592
## bmi           -898.8988    24604.0399
## education     64761.0950    72461.3161
## age           -8275.8056     646.5600

## [1] 0.2177982

##
## Call:
## lm(formula = modell_4, data = hoyde)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -994470 -230101  -57569   124319  2855576
##
## Coefficients:
##           Estimate Std. Error t value      Pr(>|t|)
## (Intercept) -2603720    413936  -6.290 0.000000000337 ***
## hoyde_cm      13468       2344   5.745 0.000000009591 ***
## vekt_kg       -4736       2272  -2.084   0.0372 *
## bmi           11853       6505   1.822   0.0685 .
## education     68611       1964  34.934 < 2e-16 ***
## age           -3815       2276  -1.676   0.0937 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 422300 on 6895 degrees of freedom
## (105 observations deleted due to missingness)
## Multiple R-squared:  0.1939, Adjusted R-squared:  0.1933
## F-statistic: 331.7 on 5 and 6895 DF,  p-value: < 2.2e-16

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



```
## Residuals:
##      Min       1Q   Median       3Q      Max
## -844832 -243051  -90336  125765 2664145
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   -2296981.00   3061743.05  -0.750   0.453
## sexfemale      364907.00   3877913.34   0.094   0.925
## hoyde_cm      14177.21    17010.78   0.833   0.405
## vekt_kg        217.78    29577.95   0.007   0.994
## I(vekt_kg^2)   -29.54      71.13  -0.415   0.678
## bmi           17853.31    96322.70   0.185   0.853
## I(bmi^2)       -64.14     767.00  -0.084   0.933
## sexfemale:hoyde_cm -31.35    22460.23  -0.001   0.999
## sexfemale:vekt_kg -18273.48   40602.63  -0.450   0.653
## sexfemale:I(vekt_kg^2) 67.20     105.64   0.636   0.525
## sexfemale:bmi     25515.27  120895.62   0.211   0.833
## sexfemale:I(bmi^2)   -193.30     937.66  -0.206   0.837
##
## Residual standard error: 455700 on 6899 degrees of freedom
## (95 observations deleted due to missingness)
## Multiple R-squared:  0.06166, Adjusted R-squared:  0.06016
## F-statistic: 41.21 on 11 and 6899 DF, p-value: < 2.2e-16
```

## Residualene

I den endelige modellen legges variablene *education*, *married*, *sex* og *afqt* til i tillegg. Variabelen *age* legges ikke til fordi aldersspennet er bare mellom 9 år, noe som ikke er særlig stor variasjon. Fullstendige modellen blir dermed seende slik ut:

```
##
## Call:
## lm(formula = modell_f, data = hoyde)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -975293 -210118  -43707  125388 2758733
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -1554319     404900  -3.839 0.000125 ***
## hoyde_cm      7114       2387   2.980 0.002889 **
## vekt_kg       -6529       2265  -2.883 0.003956 **
## bmi          17534       6471   2.710 0.006755 **
## education    49489       2419  20.456 < 2e-16 ***
## marriedTRUE   88747      10450   8.493 < 2e-16 ***
## afqt          3313        221  14.993 < 2e-16 ***
## sexfemale    -204806     14542 -14.083 < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 411000 on 6637 degrees of freedom
## (361 observations deleted due to missingness)
## Multiple R-squared:  0.255, Adjusted R-squared:  0.2542
```

```
## F-statistic: 324.5 on 7 and 6637 DF, p-value: < 2.2e-16
```

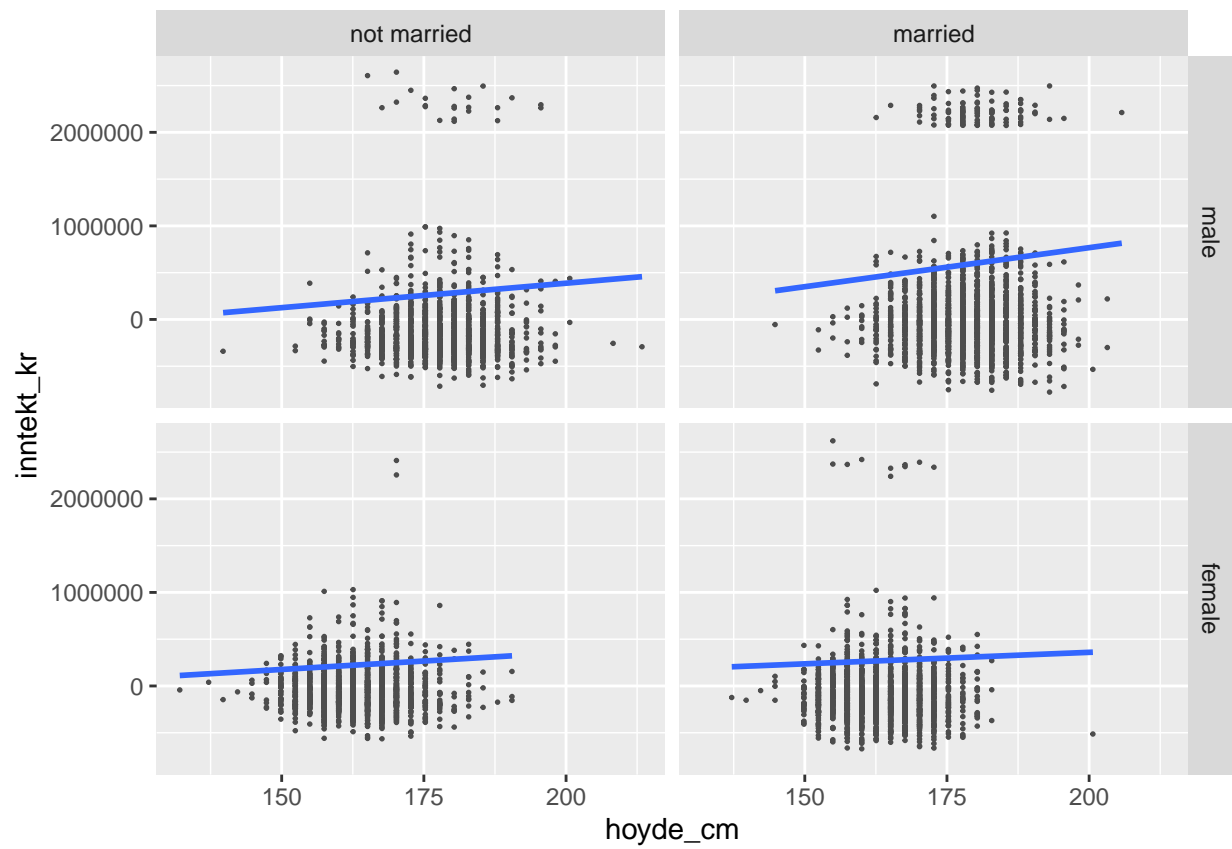
Så kjører vi den fullstendige modellen og begrenser den med å fjerne de 2% høyeste inntektene og de som har null i inntekt.

```
##
## Call:
## lm(formula = modell_f, data = hoyde_b)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -714122 -155634  -29517   117375  1102390
##
## Coefficients:
##              Estimate Std. Error t value      Pr(>|t|)
## (Intercept) -103335.1   295314.0  -0.350      0.726
## hoyde_cm      170.8     1734.2   0.098      0.922
## vekt_kg       261.5     1637.6   0.160      0.873
## bmi        -1451.0     4720.1  -0.307      0.759
## education    32413.1     1650.2  19.642 < 2e-16 ***
## marriedTRUE   46570.6     7171.1   6.494 0.0000000000918 ***
## afqt         1992.6      149.0  13.373 < 2e-16 ***
## sexfemale   -130192.0    10156.9 -12.818 < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 241200 on 4863 degrees of freedom
## (252 observations deleted due to missingness)
## Multiple R-squared:  0.2605, Adjusted R-squared:  0.2595
## F-statistic: 244.8 on 7 and 4863 DF, p-value: < 2.2e-16
```

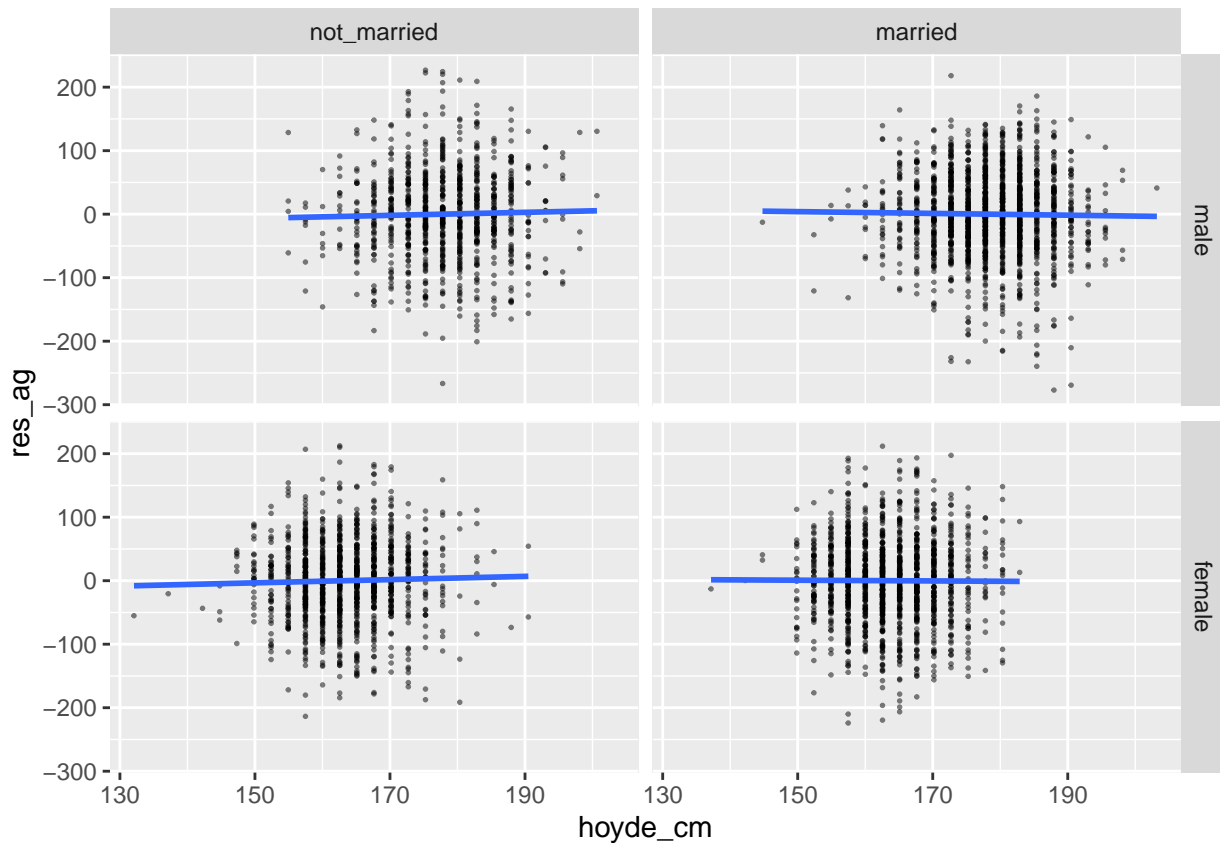
Ved å fjerne 2% høyeste inntekt og de som har null i inntekt mister *hoyde\_cm* sin signifikans.

Legger til residualene fra endelig modell fra datasettet

```
## Warning: Removed 361 rows containing missing values (geom_point).
```



Mitt forslag



Etter at man har korrigert for kjønn, utdanning, evner og sivilstatus ser det altså ikke ut til å være noen sammenheng mellom uforklart variasjon og høyde. Kan selvsagt også vises vha. regresjonsanalyse (robust se).

```
##
## t test of coefficients:
##
##           Estimate Std. Error t value Pr(>|t|)
## (Intercept) -3.903327  15.630058 -0.2497  0.8028
## hoyde_cm      0.022857   0.091866  0.2488  0.8035
##
## Call:
## lm(formula = "res_ag ~ hoyde_cm", data = .)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -277.376  -39.440    0.989   41.624  226.667
##
## Coefficients:
##           Estimate Std. Error t value Pr(>|t|)
## (Intercept) -3.90333   15.92245  -0.245  0.806
## hoyde_cm      0.02286    0.09307   0.246  0.806
##
## Residual standard error: 66.68 on 4869 degrees of freedom
## Multiple R-squared:  1.239e-05, Adjusted R-squared:  -0.000193
## F-statistic: 0.06031 on 1 and 4869 DF, p-value: 0.806
```

```
##
## Call:
## lm(formula = "res_ag ~ sex*(hoyde_cm + married)", data = .)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -277.063  -39.510    0.935   41.518  226.792
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      -1.514347   32.736223  -0.046   0.963
## sexfemale        -13.683812   46.004737  -0.297   0.766
## hoyde_cm          0.008516    0.183676   0.046   0.963
## marriedmarried    -0.006143    2.809535  -0.002   0.998
## sexfemale:hoyde_cm  0.084563    0.269771   0.313   0.754
## sexfemale:marriedmarried -0.024368    3.894700  -0.006   0.995
##
## Residual standard error: 66.7 on 4865 degrees of freedom
## Multiple R-squared:  4.606e-05, Adjusted R-squared:  -0.0009816
## F-statistic: 0.04481 on 5 and 4865 DF, p-value: 0.9988
```

Ser altså ikke ut til å være noen sammenheng for noen av gruppene.

## Konklusjon

På interaksjonsmodellen var ingen variabler signifikante. Det kan da tenkes at kjønn ikke har noe å si for *inntekt\_kr*.

Variabelen *hoyde\_cm* blir usignifikant med det begrensede datasettet ved å fjerne de 2% høyeste inntektene og de som har null i inntekt.

Vi kan dermed **ikke** påstå at det er høyde som bestemmer inntekt.

Resultatene viser helt klart at *education*, *afqt* og *sex* er de mest signifikante variablene som påvirker inntekt.

## 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. (2004a). The Effect of Physical Height on Workplace Success and Income: Preliminary Test of a Theoretical Model. *Journal of Applied Psychology*, 89(3), 428–441. <https://doi.org/10.1037/0021-9010.89.3.428>
- Judge, T. A., og Cable, D. M. (2004b). 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>
- R Core Team. (2021). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. <https://www.R-project.org/>
- Wickham, H. (2020). *modelr: Modelling Functions that Work with the Pipe*. <https://CRAN.R-project.org/package=modelr>
- 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 1948–1996. *Perceptual and Motor Skills*, 87(1), 321–322. <https://doi.org/10.2466/pms.1998.87.1.321>

	mod_4b	mod_4b robust se
(Intercept)	-110.338 [-0.848]	-110.338 [-0.770]
sexfemale	200.614 [1.107]	200.614 [1.084]
bmi	-1.725 [-1.692]	-1.725 [-1.558]
education	64.704 *** [30.332]	64.704 *** [23.595]
age	-5.237 * [-2.182]	-5.237 * [-1.989]
marriedTRUE	217.312 *** [19.684]	217.312 *** [18.491]
sexfemale:bmi	-0.172 [-0.136]	-0.172 [-0.133]
sexfemale:education	-19.460 *** [-6.628]	-19.460 *** [-5.813]
sexfemale:age	0.284 [0.085]	0.284 [0.084]
sexfemale:marriedTRUE	-204.618 *** [-13.362]	-204.618 *** [-13.523]
N	6901	6901
R2	0.273	
logLik	-49380.605	-49380.605
AIC	98783.209	98783.210

Regresjonstabell ag: \*\*\*  $p < 0.001$ ; \*\*  $p < 0.01$ ; \*  $p < 0.05$ . T statistics in brackets.

	modell_1	modell_2	modell_3
(Intercept)	-1329893.035 *** [-14.803]	-1441826.256 *** [-15.471]	-1907488.774 *** [-4.458]
hoyde_cm	9825.866 *** [18.676]	11226.677 *** [18.696]	13971.667 *** [5.513]
vekt_kg		-1481.275 *** [-4.794]	-4197.636 [-1.709]
bmi			7833.548 [1.115]
N	7006	6911	6911
R2	0.047	0.051	0.051
logLik	-101239.507	-99887.319	-99886.697
AIC	202485.014	199782.637	199783.394

Regresjonstabell 3: \*\*\*  $p < 0.001$ ; \*\*  $p < 0.01$ ; \*  $p < 0.05$ . T statistics in brackets.

Res.Df	RSS	Df	Sum of Sq	F	Pr(>F)
6.90e+03	1.45e+15				
6.9e+03	1.43e+15	6	1.5e+13	12	1.81e-13

light	weight	age	marital	sex	education	afqt	hoyde_cm	vekt_kg	inntekt_kr	bmi	married
60	155	53	married	female	13	6.84	152	70.3	1.59e+05	30.3	TRUE
70	156	51	married	female	10	49.4	178	70.8	2.93e+05	22.4	TRUE
65	195	52	married	male	16	99.4	165	88.5	8.79e+05	32.4	TRUE
63	197	54	married	female	14	44	160	89.4	3.35e+05	34.9	TRUE
66	190	49	married	male	14	59.7	168	86.2	6.28e+05	30.7	TRUE