



BAGGRUND OG DATABESKRIVELSE

Denne opgave beskæftiger sig med en eventuel sammenhæng mellem uddannelsesniveau og personlig indkomst. Opgaven tager udgangspunkt i datamaterialet i filen `Uddanne1se-Indkomst.jmp`. Datamaterialet indeholder information om alder, køn, højest gennemførte uddannelse samt årlig bruttoindkomst (i kr.) i kalenderåret 2005 for et udsnit af den danske befolkning (for at sikre enkeltpersoners anonymitet er bruttoindkomsten “skåret af” ved 1,2 mio. kr.).

EKSEMPEL 1 (KØN OG UDDANNELSE)

- a). **Tegn** for hvert køn et søjlediagram over fordelingen af uddannelsesniveau (variablen *Uddannelse*).

Tegn for hvert uddannelsesniveau et søjlediagram over fordelingen af de to køn.

Gør rede for hvad figurerne viser om ligheder og forskelle på de to køns uddannelsesvalg.

EKSEMPEL 2 (KØN OG UDDANNELSE)

- a). **Opstil** en krydstabel for køn og uddannelsesniveau.

Beregn andelen af personer i datamaterialet med en mellemlang videregående uddannelse som højest gennemførte uddannelse.

- b). **Beregn** andelen af personer i datamaterialet, der udgøres af mænd med folkeskole som højest gennemførte uddannelse.

Beregn andelen af mænd, givet at vi kun ser på personer med folkeskole som højest gennemførte uddannelse.

Gør rede for forskellen på de to beregnede andele.

- c). **Beregn** andelen af personer, der har en mellemlang eller lang videregående uddannelse som højest gennemførte uddannelse.

Beregn blandt mændene andelen, der har en mellemlang eller lang videregående uddannelse som højest gennemførte uddannelse.



EKSEMPEL 3 (UDDANNELSE OG INDKOMST)

- Tegn** et box plot over fordelingen af bruttoindkomst for hvert uddannelsesniveau.
Gør rede for hvad de optegnede box plots indikerer om ligheder og forskelle i bruttoindkomsten blandt de forskellige uddannelsesniveauer.
- Beregn** middelværdi og standardafvigelse for bruttoindkomsten for hvert uddannelsesniveau.
Gør rede for i hvilket omfang de beregnede nøgletal indikerer de samme tendenser, som de i delspg. a) optegnede box plots.

EKSEMPEL 4 (UDDANNELSE OG ALDER)

- Tegn** et scatterplot af personernes alder mod deres bruttoindkomst.
Beregn variablenes indbyrdes korrelation og **gør rede for** mulig intuition bag såvel fortegnet på som størrelsen af den beregnede korrelation.

DATAKILDE

Datamaterialet i ovenstående eksempler er venligst stillet til rådighed af **Danmarks Statistik** (dst.dk). Virksomheden giver følgende beskrivelse af sig selv:

“Danmarks Statistik er den centrale myndighed for dansk statistik. Vores opgave er at indsamle, bearbejde og offentliggøre statistiske oplysninger om det danske samfund.”

DISCLAIMER:

Datamaterialet i nærværende eksempler må udelukkende anvendes i forbindelse med undervisningen i faget Dataanalyse på CBS. Datamaterialet er fortroligt og må ikke videredistribueres eller offentliggøres andet end overfor studerende, der følger faget Dataanalyse på CBS, hverken skriftligt, mundtligt eller på anden vis, og ej heller de resultater og indsigter, der måtte komme ud af arbejdet med datamaterialet.

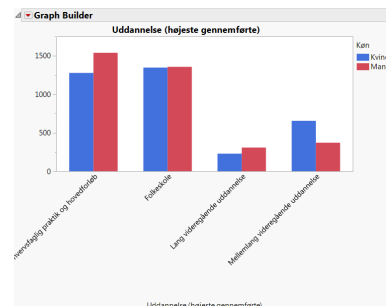
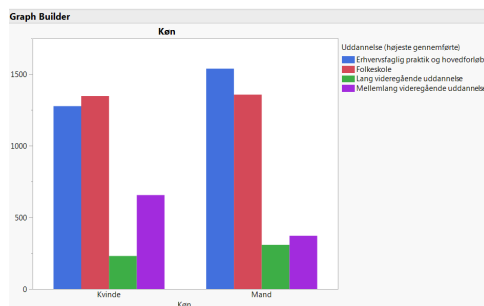


VEJLEDENDE LØSNINGER

EKSEMPEL 1 (► Løsning i JMP)

1A):

Søjlediagrammer over fordelingen af køn og uddannelsesniveau for personerne i datamaterialet:



Søjlediagrammet over fordelingen af uddannelsesniveau indenfor hvert køn (figuren til venstre) viser, at for kvinder er folkeskolen eller en erhvervsfaglig uddannelse de to hyppigste uddannelsesniveauer. De er omtrent lige hyppige, mens der kun er ca. halvt så mange, som har en mellemlang videregående uddannelse og endnu færre som har en lang videregående uddannelse. Det samme mønster gør sig gældende for mænd, med den ene undtagelse at der er nogenlunde lige mange, der har henholdsvis en mellemlang og en lang videregående uddannelse.

Søjlediagrammet over fordelingen af køn indenfor hvert uddannelsesniveau (figuren til højre) viser, at der er lidt flere mænd end kvinder med en erhvervsfaglig uddannelse, at der omvendt er flere kvinder end mænd med en mellemlang videregående uddannelse, mens kønsfordelingen er nogenlunde ligelig for så vidt angår personer med folkeskole og ligeså personer med en lang videregående uddannelse.



EKSEMPEL 2 (► Løsning i JMP)

2A):

Contingency Table					
Køn	Uddannelse (højeste gennemførte)				Total
	Erhvervsfaglig praktik og hovedforløb	Folkeskole	Lang videregående uddannelse	Mellemlang videregående uddannelse	
	Count				
	Total %				
	Col %				
	Row %				
Kvinde	1277	1347	231	656	3511
	18,02	19,01	3,26	9,26	49,54
	45,36	49,82	42,78	63,81	
	36,37	38,37	6,58	18,68	
Mand	1538	1357	309	372	3576
	21,70	19,15	4,36	5,25	50,46
	54,64	50,18	57,22	36,19	
	43,01	37,95	8,64	10,40	
Total	2815	2704	540	1028	7087
	39,72	38,15	7,62	14,51	

Krydstabellen for køn og uddannelsesniveau viser, at andelen af personer med en mellem-lang videregående uddannelse er

$$P(\text{mellemlang udd.}) = \frac{\text{antal med mellemlang udd.}}{\text{antal i datamaterialet}} = \frac{1.028}{7.087} = 14,51\%$$

2B):

Krydstabellen viser, at andelen af mænd med folkeskole som uddannelsesniveau er

$$P(\text{mænd med folkeskole}) = \frac{\text{antal mænd med folkeskole}}{\text{antal i datamaterialet}} = \frac{1.357}{7.087} = 19,15\%$$

og ligeledes at andelen af mænd, blandt personer med folkeskole som uddannelsesniveau, er

$$P(\text{mænd} | \text{folkeskole}) = \frac{\text{antal mænd med folkeskole}}{\text{antal med folkeskole}} = \frac{1.357}{2.704} = 50,18\%$$

I det omfang at datamaterialet er repræsentativt for den samlede danske befolkning viser beregningerne dermed, at 19,15% af den samlede befolkning (over 18 år) udgøres af mænd med folkeskole som uddannelsesniveau (dvs. knap hver femte). Ser vi imidlertid kun på den del af befolkningen, der har folkeskole som uddannelsesniveau, så udgør mænd 50,18% (dvs. ca. halvdelen). Forskellen på de to tal er, at ved den første andel ser vi på mænd med folkeskole som uddannelsesniveau **i forhold til den samlede befolkning**, mens vi ved den anden andel ser **i forhold til den del af befolkningen, der har folkeskole som uddannelsesniveau**.



2c):

Krydstabellen viser, at andelen af personer med en mellemlang eller lang videregående uddannelse er

$$P(\text{mellemlang eller lang udd.}) = \frac{\text{antal med mellemlang eller lang udd.}}{\text{antal i datamaterialet}} = \frac{540 + 1.028}{7.087} = 22,13\%$$

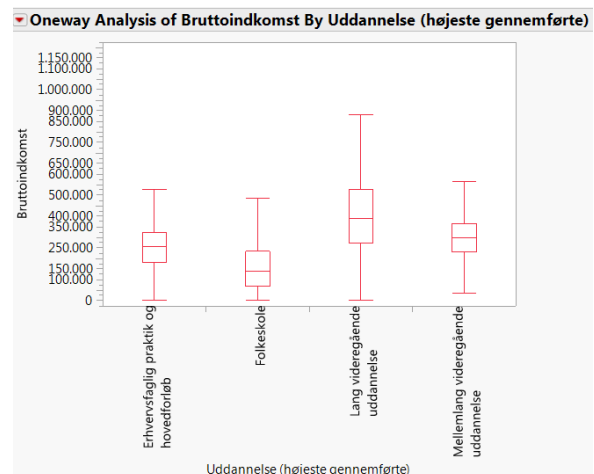
og ligeledes at blandt mændene er andelen, der har en mellemlang eller lang videregående uddannelse

$$P(\text{mellemlang eller lang udd.} | \text{mænd}) = \frac{\text{antal mænd med mellemlang eller lang udd.}}{\text{antal mænd i datamateriale}} = \frac{309 + 372}{3.576} = 19,04\%$$

Eksempel 3 (► Løsning i JMP)

3a):

Box plots af fordelingen af bruttoindkomsten indenfor hvert uddannelsesniveau viser, at i gennemsnit er indkomsten lavest for personer med folkeskole som højest gennemførte uddannelse. Herefter følger personer med en erhvervsfaglig uddannelse. Med en lidt højere indkomst end dem følger så personer med en mellemlang videregående uddannelse, mens personer med en lang videregående uddannelse i gennemsnit har den højeste bruttoindkomst.



Figurerne viser også, at variationen i bruttoindkomst fra person til person er nogenlunde den samme for alle uddannelsesniveauer med undtagelse af personer med en lang videregående uddannelse, hvor variationen er noget større.

**3B):**

Middelværdi og standardafvigelse af bruttoindkomsten indenfor hvert uddannelsesniveau viser de samme tendenser som i box

Level	Number	Mean	Std Dev	Std Err Mean	Lower 95%	Upper 95%
Erhvervsfaglig praktik og hovedforløb	2815	268350,12	136719,91	2576,8704	263297,37	273402,87
Folkeskole	2704	163173,63	132850,1	2554,8096	158164,06	168183,21
Lang videregående uddannelse	540	428217,75	245016,45	10543,829	407505,72	448929,79
Mellemlang videregående uddannelse	1028	318371,04	161982,81	5052,1051	308457,41	328284,67

plots i delspg. a), nemlig at det gennemsnitlige indkomstniveau (målt ved middelværdien) er lavest for personer med folkeskole som højest uddannelse, stigende til personer med en erhvervsfaglig uddannelse og sluttende med personer med en mellemlang og endelig en lang videregående uddannelse. Ligeledes viser tallene, at variationen i bruttoindkomst fra person til person (målt ved standardafvigelsen) er stort set den samme for alle uddannelsesniveauer med undtagelse af personer med en lang videregående uddannelse, hvor der alt andet lige er noget større indkomstforskelle fra person til person.

**Eksempel 4** (► *Løsning i JMP*)**4A):**

Scatterplottet af alder mod bruttoindkomst indikerer en vis positiv sammenhæng mellem bruttoindkomst og alder, hvilket betyder, at jo ældre man er, desto højere bruttoindkomst har man alt andet lige.

Figuren viser samtidig også, at der er stor spredning i observationerne (punkterne i scatterplottet er langt fra at ligge på en ret linje).

Begge dele understøttes af den beregnede korrelation mellem variablene på 0,2127, som er positiv og dermed ligeledes indikerer en positiv sammenhæng mellem alder og bruttoindkomst, men som samtidig ikke er større (den er langt fra at være tæt på +1) end, at den også indikerer, at sammenhængen mellem alder og indkomst ikke er overvældende. M.a.o.: Der er mange andre ting end alder, der influerer på størrelsen af den personlige indkomst.

