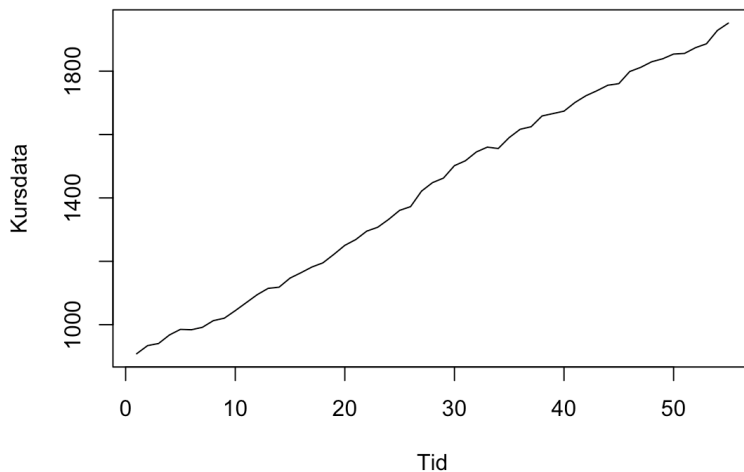


Eksamen i kvantitativ metode i den finansielle sektor

Spørgsmål 1

De historiske data for kursudviklingen plottes ind i R. Nedenfor ses en grafisk fremstilling af den historiske kursudvikling. Der ses en tydelig positiv udvikling i kursen over tid:



Ved at benytte kommandoen: `auto.arima` finder jeg frem til hvilken arima model der er tale om:

```
Series: ARIMA  
ARIMA(0,1,0) with drift
```

```
Coefficients:
```

```
drift  
19.3231
```

```
s.e. 1.4870
```

```
sigma^2 estimated as 121.7: log likelihood=-205.75  
AIC=415.5 AICc=415.73 BIC=419.48
```

R fortæller, at der altså tale om en ARIMA(0,1,0) with a drift. Denne model kan skrives som:

$$\hat{Y}_t = Y_{t-1} + \mu$$

Vi indsætter μ , givet af R. Modellen forudsiger at aktiekursen stiger med 19,3231 fra periode til periode.

$$\hat{Y}_t = Y_{t-1} + 19,3231$$

Herudfra kan jeg nu forudsige aktiekursen for de kommende 3 perioder, her kaldet tidsperiode: 56, 57, 58. Dette kan jeg nu selv beregne således:

Tidsperiode	Beregning	Kurs
55		1951,66
56	$\hat{Y}_t = 1951,66 + 19,3231$	1970,98
57	$\hat{Y}_t = 1970,98 + 19,3231$	1990,31
58	$\hat{Y}_t = 1990,31 + 19,3231$	2009,63

Mine beregninger kan også bekræftes af R ved kommandoen `forecast(auto.arima)`, hvor nedenstående output fremkommer. Nedenstående viser også den forventede aktiekurs i de kommende perioder.

	Point Forecast	Lo 80	Hi 80	Lo 95	Hi 95
56	1970.983	1956.847	1985.119	1949.364	1992.602
57	1990.306	1970.315	2010.298	1959.732	2020.880
58	2009.629	1985.145	2034.114	1972.184	2047.075
59	2028.953	2000.681	2057.224	1985.714	2072.191
60	2048.276	2016.667	2079.885	1999.934	2096.617
61	2067.599	2032.973	2102.225	2014.643	2120.555
62	2086.922	2049.522	2124.322	2029.723	2144.121
63	2106.245	2066.263	2146.228	2045.097	2167.393
64	2125.568	2083.160	2167.976	2060.711	2190.426
65	2144.891	2100.190	2189.593	2076.526	2213.257

Yderligere viser R, at:

Kursen vil med 95% sandsynlighed ligge mellem 1949,36-1992,60 i periode 1 (56).

Kursen vil med 95% sandsynlighed ligge mellem 1959,73-2020,88 i periode 2 (57)

Kursen vil med et 95% sandsynlighed ligge mellem 1972,18-2047,075 i periode 3 (58).

Her er intervallet større, da det er forbundet med en større usikkerhed at forudsige kursen længere ude i fremtiden.

Spørgsmål 2

Til beregning af om gennemsnitsalderen er ens i de 5 regioner anvendes en ANOVA-analyse. ANOVA anvendes til at sammenligne middelværdierne for variablene, når vi har en kvalitativ variabel = regioner og en kvantitativ variabel = alder.

ANOVA/ANAVA	
Forudsætninger	Populationerne har samme varians, dette er undersøgt vha. Bartlett's test for varianshomogenitet. Da p-værdien 69,28% er større end signifikansniveauet 5% kan vi ikke forkaste nulhypotesen, og der er varianshomogenitet, se nedenfor. Populationerne er normalfordelte dette undersøges vha. normalfraktildiagram. Punkter følger tilnærmelsesvis den rette linje, den indsamlede data 'alder', kan derfor anse for at være normalfordelt, se nedenfor.

Hypotese	H_0 : Alle middelværdier er ens. H_1 : Ikke alle middelværdier er ens.
Teststørrelse	F teststørrelse 0,8862
P-værdi/signifikanssandsynlighed	47,18%
Signifikansniveau	5%
Teknisk konklusion	Da forudsætningerne for anvendelse af ANOVA er opfyldt, kan vi nu se på resultatet heraf. Da p-værdien 47,18% er klart større end signifikansniveauet på 5%, kan vi ikke forkaste nulhypotesen. Den store p-værdi gør, at vi er ret sikre på vores konklusion.
Ikke-teknisk konklusion	Vi kan konkludere, at gennemsnitsalderen i de 5 regioner er ens.

Bartlett's test for varianshomogenitet

Teststørrelse

2,2340

Signifikansniveau

5%

p-værdi

69,2804%

Kritisk værdi

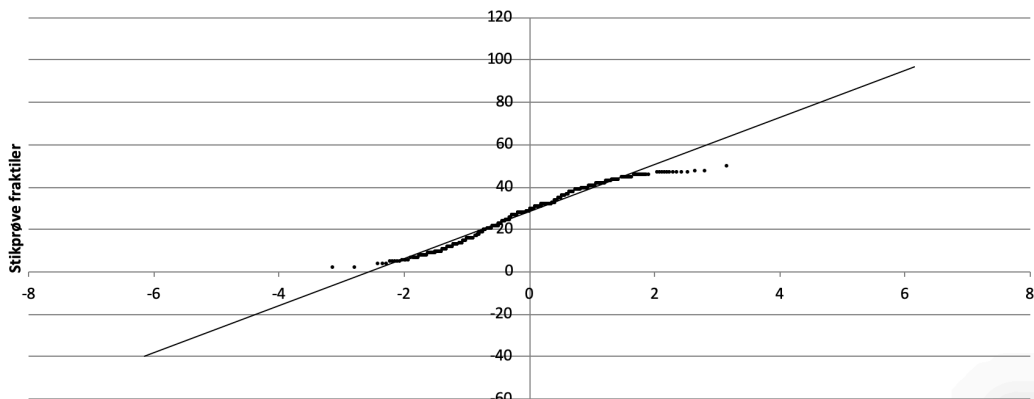
9,4877

H_0 : Alle varianser er ens, der er varianshomogenitet

H_A : Ikke alle varianser er ens, der er ikke varianshomogenitet

Vi benytter Chi i anden fordelingen med 4 frihedsgrader. Da Chi i anden teststørrelsen 2,23, er mindre end den kritiske værdi 9,49, og p-værdien 0,6928 er større end signifikansniveauet 0,05, kan vi ikke afvise nulhypotesen om varianshomogenitet.

Normalfraktildiagram



ANOVA-analyse:

F teststørrelse

Signifikans niveau

128,3906104

113,7806691

0,8862

5,00%

P værdi

Critical value

47,18%

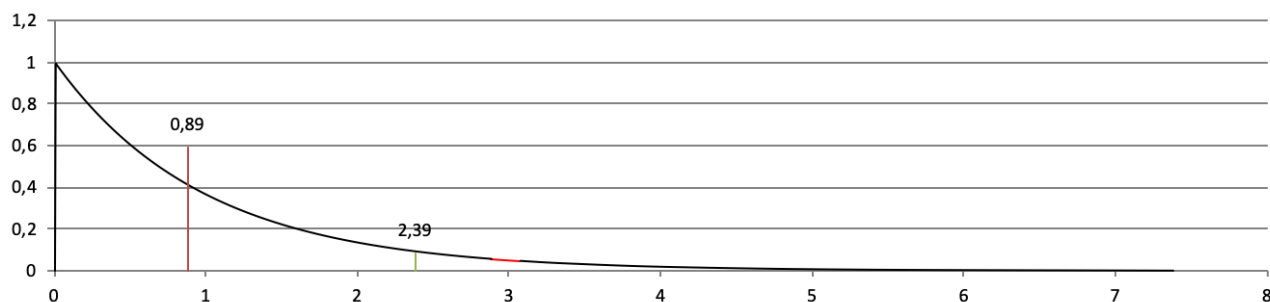
2,386934691

H_0 : Alle middelværdier er ens

H_A : Ikke alle middelværdier er ens

F(4,594). Da F teststørrelsen 0,8862, er mindre end den kritiske værdi 2,3869, og p-værdien 0,4718 er større end signifikansniveauet 0,05, kan vi ikke afvise nulhypotesen om at alle middelværdier er identiske

F(4,594). Da F teststørrelsen 0,8862, er mindre end den kritiske værdi 2,3869, og p-værdien 0,4718 er større end signifikansniveauet 0,05, kan vi ikke afvise nulhypotesen om at alle middelværdier er identiske



Spørgsmål 3:

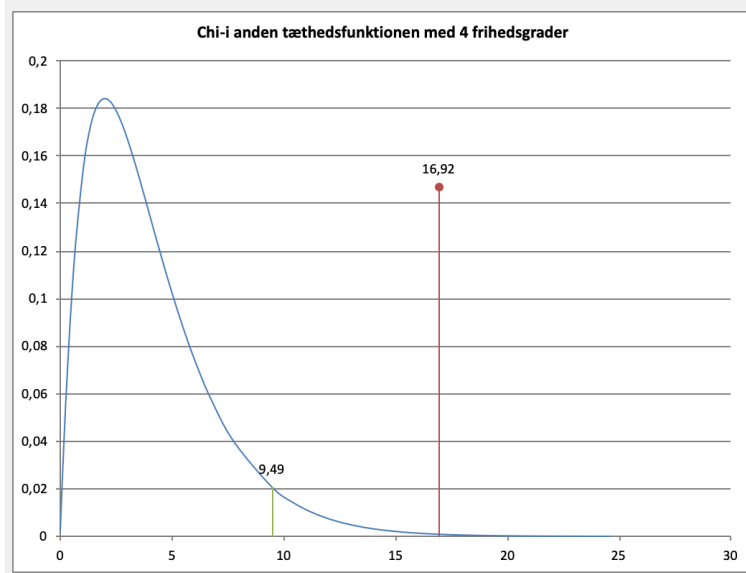
Til analysen om der er sammenhæng mellem region og køn anvendes Chi-i-anden test, da denne anvendes til at undersøge om der er uafhængighed/ afhængighed mellem 2 kvalitative variable: region og køn.

Chi i anden test																			
Forudsætninger	De forventede celleværdier skal være større end 5, for at sikre kvaliteten af testen. Alle celler opfylder dette kriterie: <table><tr><th colspan="3">Forventet</th></tr><tr><td>Række 1</td><td>90,39065109</td><td>97,6093489</td></tr><tr><td>Række 2</td><td>63,94657763</td><td>69,0534224</td></tr><tr><td>Række 3</td><td>33,17529215</td><td>35,8247078</td></tr><tr><td>Række 4</td><td>49,52253756</td><td>53,4774624</td></tr><tr><td>Række 5</td><td>50,96494157</td><td>55,0350584</td></tr></table>	Forventet			Række 1	90,39065109	97,6093489	Række 2	63,94657763	69,0534224	Række 3	33,17529215	35,8247078	Række 4	49,52253756	53,4774624	Række 5	50,96494157	55,0350584
Forventet																			
Række 1	90,39065109	97,6093489																	
Række 2	63,94657763	69,0534224																	
Række 3	33,17529215	35,8247078																	
Række 4	49,52253756	53,4774624																	
Række 5	50,96494157	55,0350584																	
Hypotese	H_0 : Der er uafhængighed mellem region og køn H_1 : Der er afhængighed mellem region og køn																		
Teststørrelse	16,92																		
P-værdi/signifikanssandsynlighed	0,20%																		
Signifikansniveau	5%																		
Teknisk konklusion	Da p-værdien 0,20% er mindre end signifikansniveauet 5% vi kan forkaste nulhypotesen. Da p-værdien er mindre end 1% er vi ret sikre på vores konklusion.																		
Ikke-teknisk konklusion	Der er sammenhæng mellem region og køn. Dette ses bl.a. i Region Sjælland, hvor vi forventede der var 49 kvinder, men havde observeret 63, og det forventes at der var 53 mænd, men der var 40 i observationen. I region Nordjylland forventes andel af mænd og kvinder at være nogenlunde lige fordel med henholdsvis 33 kvinder og 35 mænd, men observationen viste, at der var 22 kvinder og 47 mænd.																		

Frihedsgrader	4	Kritisk værdi	9,488	Signifikans niveau	5,00%
Chi-teststørrelse	16,92070113				
p-værdi	0,200273582%				

H_0 : Der er uafhængighed mellem række- og søjlekriterie
 H_A : Der er afhængighed mellem række- og søjlekriterie

P-værdien/signifikanssandsynligheden 0,002 er mindre end signifikansniveauet 5%, Det betyder Chi-teststørrelsen 16,92 er større end den kritiske værdi 9,49. Vi afviser nulhypotesen, der er afhængighed mellem række- og søjlekriterie.



Chi i anden ordnede data

	Kvinde	Mand
Region Hovedstaden	90	98
Region Midtjylland	69	64
Region Nordjylland	22	47
Region Sjælland	63	40
Region Syddanmark	44	62
Række 6		
Række 7		
Række 8		
Række 9		
Række 10		
Sum	288	311
Forventet		
Række 1	90,39065109	97,6093489
Række 2	63,94657763	69,0534224
Række 3	33,17529215	35,8247078
Række 4	49,52253756	53,4774624
Række 5	50,96494157	55,0350584

Spørgsmål 4:

Til test om stikprøven er repræsentativ i forhold til befolkningen anvendes Goodness-of-fit, denne finder anvendelse når vi ønsker at sammenligne vores stikprøve med et antal forudsætninger, som er situationen her.

Goodness of fit	
Forudsætninger	De forventede celleværdier skal være større end 5, for at sikre kvaliteten af testen. Denne forudsætning er opfyldt, se nedenfor.
Hypotese	H_0 : Stikprøvefordelingen stemmer overens med populationsfordelingen fra DK statestik H_1 : Stikprøvefordelingen stemmer ikke overens med populationsfordelingen fra DKstatestik
Teststørrelse	7,52
P-værdi/signifikanssandsynlighed	11,07%
Signifikansniveau	5%
Teknisk konklusion	Da p-værdien 11,07% er større end signifikansniveauet på 5% kan vi ikke forkaste nulhypotesen.
Ikke-teknisk konklusion	Stikprøvefordelingen stemmer overens med folketallet i 2017 fra Danmarks statistik. Stikprøven er derfor repræsentativ i forhold til befolkningens fordeling på regioner.

	Antal observeret	Sandsynligheder forventet	Chi square bidrag	Forventet antal
Region Hovedstader	188	0,314398439	0,000559711	188,324665
Region Midtjylland	133	0,22687518	0,061809137	135,898233
Region Nordjylland	69	0,10216709	0,994636444	61,1980869
Region Sjælland	103	0,144822831	3,044408834	86,7488756
Region Syddanmark	106	0,21173646	3,421069512	126,83014
Frihedsgrader	4	Kritisk værdi	9,488	
Chi teststørrelse	7,522483638			
p-værdi	11,072194%			
Signifikans niveau	5,00%			

H_0 :	Datafordelingen er identisk med den angivne fordeling
H_A :	Datafordelingen er ikke identisk med den angivne fordeling
P-værdien/signifikanssandsynligheden 0,1107 er større end signifikansniveauet 5%, Det betyder at Chi-teststørrelsen 7,5225 er mindre end den kritiske værdi 9,49. Vi kan ikke afvise nulhypotesen om at datafordelingen er identisk med den angivne fordeling.	

Spørgsmål 5

Virksomheden bør benytte sig af en klyngeanalyse, da denne strukturanalyse anvendes til at inddele respondenter efter samme karakteristika.

Spørgsmål 6

Nedenfor ses respondenterne fra Nordjyllands svar inddelt i klyngeplot i 3 klynger. De er inddelt efter hierarkisk klyngedannelse. Denne metode er at foretrække, når vi ikke har meget store datasæt, som i opgaven her.

Klyngeplottet er udarbejdet ved hjælp af R og Kmeans kommandoen.

