

Udvidet Statistik for de finansielle uddannelser.

*Thomas Petersen*

*2018-07-17*



# Contents

<b>Stat2</b>	<b>5</b>
<b>1 Datasæt og data</b>	<b>7</b>
<b>2 Dette er anden nummerede fil</b>	<b>9</b>
<b>3 Methods</b>	<b>11</b>
<b>4 Applications</b>	<b>13</b>
4.1 Example one . . . . .	13
4.2 Example two . . . . .	13
<b>5 Indledning</b>	<b>15</b>
<b>6 Applications tilføjet</b>	<b>17</b>
6.1 Example one . . . . .	17
6.2 Example two . . . . .	17
<b>7 Introduction</b>	<b>19</b>
<b>8 Datasæt og data44</b>	<b>23</b>
<b>9 Datasæt og data22</b>	<b>27</b>
<b>10 Indledning Start</b>	<b>31</b>
10.1 Does Ruby work with <b>knitr</b> ? . . . . .	31
10.2 js . . . . .	31
10.3 Bash . . . . .	31
10.4 A normal R code chunk . . . . .	32
10.5 Modify an R variable . . . . .	32
10.6 A Python chunk . . . . .	33
10.7 Modify a Python variable . . . . .	33
10.8 Python graphics . . . . .	33
10.9 Opgaver . . . . .	35
10.10 Quarterly Results . . . . .	35
10.11 Nyt afsnit efter tabbed . . . . .	35



# Stat2

Login eller opret klik her

.

Glemt? Vis

klik

Husk mig

.

Min profil

Opret

Log Out

Javascript Required

For testing.

You must have JavaScript enabled in order to log in.

# Indledning

Dette er undervisningsmateriale og opgaver til faget statistik, for erhvervsakademierne.

Orley Ashenfelter en Princeton økonom udviklede i 1980'erne en statistik model til forudsigelse af vinpriser baseret på nedbør, solskinstimer og andre klimadata. Hele den etablerede vinverden var i oprør, ved en præsentation i Christie's vinafdeling, blev han buhet ud. Robert Parker den verdenskendte vinkender udtalte "Det svarer til en filmanmelder der ikke ser filmen, men udelukkende baserer sin anmeldelse på instruktøren og skuespilleren". Orley udtalte, lang tid før det var muligt for vinseksperterne, at 1989 Bordeaux ville blive århundredets vin, uanset den kun havde ligget 3 måneder på fade. Flere analyser har siden vist Orleys model er langt mere præcis eksperterne. Meget få vinkendere har anerkendt kvaliteten af Orleys model, men deres forecasts ligger nu langt tættere på modellens forudsigelser.

Bogen er opbygget med en del praktiske eksempler.

Der er i nogle afsnit knapper med spørgsmål og svar, man kan klikke på disse og se om man kan nå frem til de rigtige løsninger.

Bogen er bygget op så kapitlerne beskriver fanerne i Freestat programmet. Man kan se og hente excelfiler direkte ved at klikke på links.

I alle brancher i den finansielle sektor spiller statistik en rolle.

Bankerne sammensætter investeringsporteføljer, der minimerer risikoen (variansen), ved aktiver der har lav eller negativ samvariation (kovarians). Cykliske aktier som FL Smidth har fx. lav samvariation med en ikke cyklisk aktie som Novo.

Forsikringsselskaberne beregner præmier for forsikringstageren, baseret på statistiske sandsynligheder for at en hændelse indtræffer. Modellerne kan være meget specifikke, en indboforsikring kan fx. være baseret på ikke bare postnummer, boligform, men også etage.

Finansielle virksomheder underlagt finanstilsynet, bruger modeller til beregning af risiko baseret på statistisk analyse.

Mægleren beregner udbudspriser, udfra en multipel lineær regressionsmodel, der indeholder variable som størrelse, energimærke, tagtype etc.

#### 0.0.0.1 Freestat

Man kan få beregnet deskriptorerne i et utal af programmer heriblandt Freestat et gratis program, der kan hentes ved at klikke på billedet. Freestat kan gennemføre de mest almindelige statistiske analyser.

Der findes opgaver quizzes og yderligere resourcer på [www.edutest.dk](http://www.edutest.dk)

Min gode ven Benjamin Tejlbjerg har lavet en super hjemmeside med gymnasie matematik og statistik <http://www.mathhx.dk>. Siden er gratis og god til at genopfriske basisbegreber indenfor statistik, vi kommer ikke i dybden med disse begreber her.

Dette er 1. udgave af bogen, der tages forbehold for tryk og tastefejl, men alle fejl eller uklarheder i måtte finde rettes med fluks. Forslag til forbedringer modtages med kyshånd.

For at få et overblik over hypoteserne for de forskellige tests kan man hente et hypotese mindmap ved at klikke på billedet til højre: \_\_\_\_\_

*Noterne er kun til personligt brug. Alle rettigheder forbeholdes. Fotografisk eller anden gengivelse af eller kopiering eller anden udnyttelse, er uden forfatterens skriftlige samtykke forbudt ifølge dansk lov om ophavsret.*

# Chapter 1

## Datasæt og data

### 1.0.1 Uni- bi- og multivariate datasæt

Datasæt er sæt af en eller flere variable:

- Univariate datasæt fx tider ved marathonløb
- Bivariate datasæt fx tider ved marathonløb og køn
- Multivariate datasæt fx tider ved marathonløb, køn, alder, medlem af sports klub

### 1.0.2 Kvalitative variable

Kvalitative variable er data vi ikke kan måle eller tælle. De antager værdier i form af navne eller labels:

- Kæledyr: kat, hund, marsvin
- Køn: mand, kvinde
- Favorit app: Angry Birds, Messenger, Audible, Tinder

### 1.0.3 Kvantitative variable

Kvantitative variable er målbare numeriske variable, vi deler disse op i *kontinuerte* og *diskrete* variable

#### 1.0.3.1 Diskrete variable

Diskrete variable er fx.

- Antal biler der passerer en bro observeret over flere dage.
- Dagsproduktionen af chokoladefrøer på Toms.
- Antal personer der har iPhones
- Antallet af indbyggere i en by

#### 1.0.3.2 Kontinuerte variable

Kontinuerte variable er fx.

- Antal ml. indhold i shampoo flasker
- Aktiekurser for Intel
- Vægten på værnepligtige
- Højden på studerende

### 1.0.4 Skaltyper

Vi kan ydermere inddele variable efter skalatype hvor lavere betyder mindst restriktiv.

1. Nominalskala, bruges til at måle kvalitative data (er der kun 2 mulige udfald kaldes variabelen specielt binær eller dikotom), fx.
  - Køn Mand Kvinde
  - Styresystem: IOS Android Windows Symbian Andet
  - Race: Europæisk, Afrikansk, Asiatisk Andet
2. Ordinalskala inddeler data efter en rangordning
  - Karakterer på 7 trins skalaen -3 00 02...
  - Moodys credit ratings Aaa Aa A Baa Ba B Caa Ca C
  - Tilfredshed meget utilfreds, noget utilfreds, nogenlunde tilfreds, meget tilfreds
3. Intervalskala man kan sammenligne afstande og forskelle, men der er intet meningsfuldt nulpunkt. Nul for en intervalskala variabel betyder således ikke fravær af den målte størrelse. Nul grader celsius betyder altså ikke fravær af temperatur (det absolutte nulpunkt 0 Kelvin, hvor alle molekyler og atomer er i grundtilstanden). En IQ på 0 betyder ikke fravær af intelligens.
  - Temperatur målt i Celsius
  - Temperatur målt i Fahrenheit
  - PH
  - IQ
4. Ratioskala
  - Beløb i lommen
  - Højde på studerende
  - Hastighed af biler ved vejkryds
  - Indhold i Coca Cola flasker

“Statistics are used much like a drunk uses a lamppost: for support, not illumination.”

- Vin Scully

Interval- og ratioskalaer omtales som numeriske eller kontinuerte skalaer, disse er knyttet til kvantitative variable.

Nominal- og ordinalskalaer omtales ofte som kategorisk eller faktor, disse er knyttet til kvalitative variable.

En stikprøve af skalatype ratio kan fx. reduceres til ordinal, eller nominal. Temperatur målt i celsius kan fx. omskrives til en ordinal variabel: koldt normalt varmt, eller en nominal variabel: ekstrem temperatur eller normal temperatur.

Kategoriske skalaer kan yderligere reduceres til en dikotom skala, ved at sammenlægge kategorierne, til man kun har 2 kategorier.

Det er vanskeligere at ændre en nominal- ordinal- eller ratioskala til en intervalskala. At ændre variabelen nominalskala variabelen køn til ordinal giver fx. ikke mening.



## Chapter 2

# Dette er anden nummerede fil

Here is a review of existing methods.



## Chapter 3

# Methods

We describe our methods in this chapter.



## Chapter 4

# Applications

Sentry Page Protection

Please Wait...

Some *significant* applications are demonstrated in this chapter.

### 4.1 Example one

### 4.2 Example two



# Chapter 5

## Indledning

Dette er noter og opgaver til faget statistik, for erhvervsakademierne.

Orley Ashenfelter en Princeton økonom udviklede i 1980'erne en statistik model til forudsigelse af vinpriser baseret på nedbør, solskinstimer og andre klimadata. Hele den etablerede vinverden var i oprør, ved en præsentation i Christie's vinafdeling, blev han buhet ud. Robert Parker den verdenskendte vinkender udtalte "Det svarer til en filmanmelder der ikke ser filmen, men udelukkende baserer sin anmeldelse på instruktøren og skuespilleren". Orley udtalte, lang tid før det var muligt for vinseksperterne, at 1989 Bordeaux ville blive århundredets vin, uanset den kun havde ligget 3 måneder på fade. Flere analyser har siden vist Orleys model er langt mere præcis eksperterne. Meget få vinkendere har anerkendt kvaliteten af Orleys model, men deres forecasts ligger nu langt tættere på modellens forudsigelser.

Bogen er opbygget med en del praktiske eksempler.

Der er i nogle afsnit knapper med spørgsmål og svar, man kan klikke på disse og se om man kan nå frem til de rigtige løsninger.

Bogen er bygget op så kapitlerne beskriver fanerne i Freestat programmet. Man kan se og hente excelfiler direkte ved at klikke på links.

I alle brancher i den finansielle sektor spiller statistik en rolle.

Bankerne sammensætter investeringsporteføljer, der minimerer risikoen (variansen), ved aktiver der har lav eller negativ samvariation (kovarians). Cykliske aktier som FL Smidth har fx. lav samvariation med en ikke cyklisk aktie som Novo.

Forsikringsselskaberne beregner præmier for forsikringstageren, baseret på statistiske sandsynligheder for at en hændelse indtræffer. Modellerne kan være meget specifikke, en indboforsikring kan fx. være baseret på ikke bare postnummer, boligform, men også etage.

Finansielle virksomheder underlagt finanstilsynet, bruger modeller til beregning af risiko baseret på statistisk analyse.

Mægleren beregner udbudspriser, ud fra en multipel lineær regressionsmodel, der indeholder variable som størrelse, energimærke, tagtype etc.

### 5.0.0.1 Freestat

Man kan få beregnet deskriptorerne i et utal af programmer heriblandt Freestat et gratis program, der kan hentes ved at klikke på billedet. Freestat kan gennemføre de mest almindelige statistiske analyser.

Der findes opgaver quizzes og yderligere resourcer på [www.edutest.dk](http://www.edutest.dk)

Min gode ven Benjamin Tejlbjerg har lavet en super hjemmeside med gymnasie matematik og statistik <http://www.mathhx.dk>. Siden er gratis og god til at genopfriske basisbegreber indenfor statistik, vi kommer ikke i dybden med disse begreber her.

Dette er 1. udgave af bogen, der tages forbehold for tryk og tastefejl, men alle fejl eller uklarheder i måtte finde rettes med fluks. Forslag til forbedringer modtages med kyshånd.

For at få et overblik over hypoteserne for de forskellige tests kan man hente et hypotese mindmap ved at klikke på billedet til højre: \_\_\_\_\_

*Noterne er kun til personligt brug. Alle rettigheder forbeholdes. Fotografisk eller anden gengivelse af eller kopiering eller anden udnyttelse, er uden forfatterens skriftlige samtykke forbudt ifølge dansk lov om ophavsret.*



## Chapter 6

# Applications tilføjet

Some *significant* applications are demonstrated in this chapter.

### 6.1 Example one

Her er lidt tekst

### 6.2 Example two

Og her

#### 6.2.1 Hvad med denne linje?



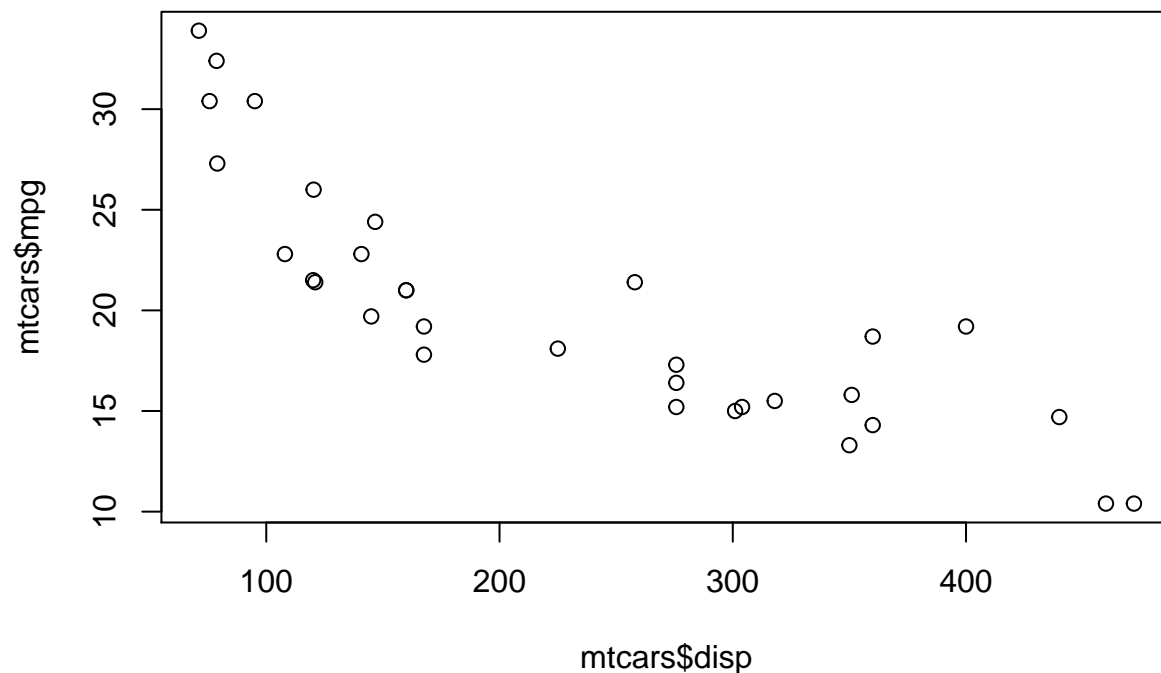
## Chapter 7

# Introduction

Subscribe

Show/Hide

```
plot(mtcars$dis, mtcars$mpg)
```



Dette er noter og opgaver til faget statistik, for erhvervsakademier og grundlæggende statistik på videregående uddannelser.

Hello world  $\alpha$

Dette er noter og opgaver til faget statistik, for erhvervsakademier og grundlæggende statistik på videregående uddannelser.

##	mpg	cyl	dis	hp	drat	wt	qsec	vs	am	gear	carb
## Mazda RX4	21.0	6	160.0	110	3.90	2.620	16.46	0	1	4	4
## Mazda RX4 Wag	21.0	6	160.0	110	3.90	2.875	17.02	0	1	4	4
## Datsun 710	22.8	4	108.0	93	3.85	2.320	18.61	1	1	4	1
## Hornet 4 Drive	21.4	6	258.0	110	3.08	3.215	19.44	1	0	3	1
## Hornet Sportabout	18.7	8	360.0	175	3.15	3.440	17.02	0	0	3	2

## Valiant	18.1	6	225.0	105	2.76	3.460	20.22	1	0	3	1
## Duster 360	14.3	8	360.0	245	3.21	3.570	15.84	0	0	3	4
## Merc 240D	24.4	4	146.7	62	3.69	3.190	20.00	1	0	4	2
## Merc 230	22.8	4	140.8	95	3.92	3.150	22.90	1	0	4	2
## Merc 280	19.2	6	167.6	123	3.92	3.440	18.30	1	0	4	4
## Merc 280C	17.8	6	167.6	123	3.92	3.440	18.90	1	0	4	4
## Merc 450SE	16.4	8	275.8	180	3.07	4.070	17.40	0	0	3	3
## Merc 450SL	17.3	8	275.8	180	3.07	3.730	17.60	0	0	3	3
## Merc 450SLC	15.2	8	275.8	180	3.07	3.780	18.00	0	0	3	3
## Cadillac Fleetwood	10.4	8	472.0	205	2.93	5.250	17.98	0	0	3	4
## Lincoln Continental	10.4	8	460.0	215	3.00	5.424	17.82	0	0	3	4
## Chrysler Imperial	14.7	8	440.0	230	3.23	5.345	17.42	0	0	3	4
## Fiat 128	32.4	4	78.7	66	4.08	2.200	19.47	1	1	4	1
## Honda Civic	30.4	4	75.7	52	4.93	1.615	18.52	1	1	4	2
## Toyota Corolla	33.9	4	71.1	65	4.22	1.835	19.90	1	1	4	1
## Toyota Corona	21.5	4	120.1	97	3.70	2.465	20.01	1	0	3	1
## Dodge Challenger	15.5	8	318.0	150	2.76	3.520	16.87	0	0	3	2
## AMC Javelin	15.2	8	304.0	150	3.15	3.435	17.30	0	0	3	2
## Camaro Z28	13.3	8	350.0	245	3.73	3.840	15.41	0	0	3	4
## Pontiac Firebird	19.2	8	400.0	175	3.08	3.845	17.05	0	0	3	2
## Fiat X1-9	27.3	4	79.0	66	4.08	1.935	18.90	1	1	4	1
## Porsche 914-2	26.0	4	120.3	91	4.43	2.140	16.70	0	1	5	2
## Lotus Europa	30.4	4	95.1	113	3.77	1.513	16.90	1	1	5	2
## Ford Pantera L	15.8	8	351.0	264	4.22	3.170	14.50	0	1	5	4
## Ferrari Dino	19.7	6	145.0	175	3.62	2.770	15.50	0	1	5	6
## Maserati Bora	15.0	8	301.0	335	3.54	3.570	14.60	0	1	5	8
## Volvo 142E	21.4	4	121.0	109	4.11	2.780	18.60	1	1	4	2

# create an anchor

create an anchor2

Orley Ashenfelter en Princeton økonom udviklede i 1980'erne en statistik model til forudsigelse af vinpriser baseret på nedbør, solskinstimer og andre klimadata. Hele den etablerede vinverden var i oprør, ved en præsentation i Christie's vinafdeling, blev han buhet ud. Robert Parker den verdenskendte vinkender udtalte "Det svarer til en filmanmelder der ikke ser filmen, men udelukkende baserer sin anmeldelse på instruktøren og skuespilleren". Orley udtalte, lang tid før det var muligt for vinseksperterne, at 1989 Bordeaux ville blive århundredets vin, uanset den kun havde ligget 3 måneder på fade. Flere analyser har siden vist Orleys model er langt mere præcis eksperterne. Meget få vinkendere har anerkendt kvaliteten af Orleys model, men deres forecasts ligger nu langt tættere på modellens forudsigelser.

Bogen er opbygget med en del praktiske eksempler.

Der er i nogle afsnit knapper med spørgsmål og svar, man kan klikke på disse og se om man kan nå frem til de rigtige løsninger.

Bogen er bygget op så kapitlerne beskriver fanerne i Freestat programmet. Man kan se og hente excelfiler direkte ved at klikke på links.

### 7.0.0.1 Freestat 44

Man kan få beregnet deskriptorerne i et utal af programmer heriblandt Freestat, der kan hentes herunder. Freestat kan gennemføre de mest almindelige statistiske analyser, der er flere forskellige versioner.

block quote

Hent Freestat til finansøkonomer her

Hent Freestat fuld version her

Take me to pookie And the destination anchor:

Der findes opgaver quizzes og yderligere resourcer på [www.edutest.dk](http://www.edutest.dk)

Min gode ven Benjamin Tejlbjerg har lavet en super hjemmeside med gymnasie matematik og statistik <http://www.mathhx.dk>. Siden er gratis og god til at genopfriske basisbegreber indenfor statistik, vi kommer ikke i dybden med disse begreber her.

For at få et overblik over hypoteserne for de forskellige tests kan man hente et hypotese mindmap, der er mindmaps til flere forskellige uddannelser under materialer. Her er hypoteseoversigten for finansøkonomer: CPH Finansøkonom statistik mindmap

Dette er 1. udgave af bogen, der tages forbehold for tryk og tastefejl, men alle fejl eller uklarheder i måtte finde rettes med fluks. Forslag til forbedringer modtages med kyshånd.

Der er en samlet oversigt over materialerne, der bruges i denne bog, disse ligger i som linket materialer i menuen for oven.

## 7.0.1 Hvad bruger man statistik til.44

I alle brancher i den finansielle sektor spiller statistik en rolle.

```
library(DT)
datatable(mtcars)
```

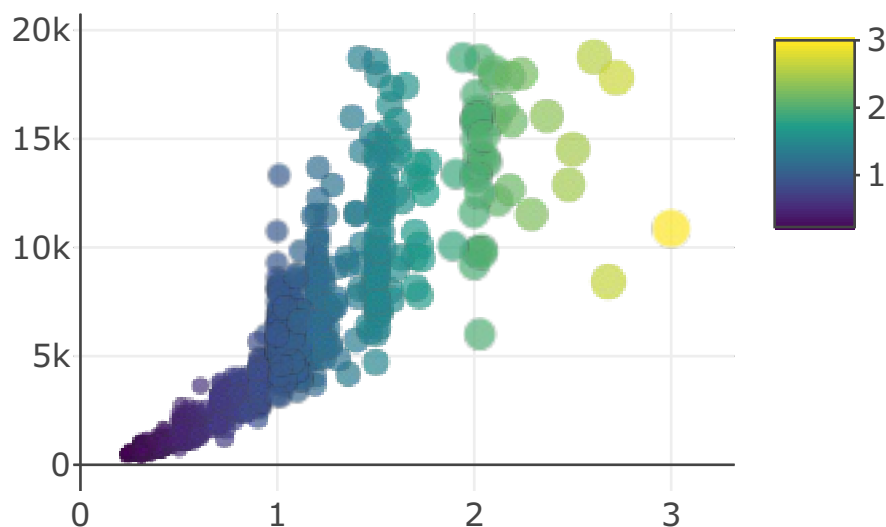
Show  entries

Search:

	mpg	cyl	disp	hp	drat	wt	qsec	vs	am	gear	carb
Mazda RX4	21	6	160	110	3.9	2.62	16.46	0	1	4	4
Mazda RX4 Wag	21	6	160	110	3.9	2.875	17.02	0	1	4	4
Datsun 710	22.8	4	108	93	3.85	2.32	18.61	1	1	4	1
Hornet 4 Drive	21.4	6	258	110	3.08	3.215	19.44	1	0	3	1
Hornet Sportabout	18.7	8	360	175	3.15	3.44	17.02	0	0	3	2
Valiant	18.1	6	225	105	2.76	3.46	20.22	1	0	3	1
Duster 360	14.3	8	360	245	3.21	3.57	15.84	0	0	3	4
Merc 240D	24.4	4	146.7	62	3.69	3.19	20	1	0	4	2
Merc 230	22.8	4	140.8	95	3.92	3.15	22.9	1	0	4	2
Merc 280	19.2	6	167.6	123	3.92	3.44	18.3	1	0	4	4

Showing 1 to 10 of 32 entries

Previous  2 3 4 Next



Bankerne sammensætter investeringsporteføljer, der minimerer risikoen (variansen), ved aktiver der har lav eller negativ samvariation (kovarians). Cykliske aktier som FL Smidth har fx. lav samvariation med en ikke cyklisk aktie som Novo.

Forsikringsselskaberne beregner præmier for forsikringstageren, baseret på statistiske sandsynligheder for at en hændelse indtræffer. Modellerne kan være meget specifikke, en indboforsikring kan fx. være baseret på ikke bare postnummer, boligform, men også etage, uddannelsesniveau etc.

Finansielle virksomheder underlagt finanstilsynet, bruger modeller til beregning af risiko baseret på statistisk analyse.

Mægleren beregner udbudspriser, ud fra en multipel lineær regressionsmodel, der indeholder variable som størrelse, energimærke, tagtype etc.

---

*Noterne er kun til personligt brug. Alle rettigheder forbeholdes. Fotografisk eller anden gengivelse af eller kopiering eller anden udnyttelse, er uden forfatterens skriftlige samtykke forbudt ifølge dansk lov om ophavsret.*

# Chapter 8

## Datasæt og data44

### 8.0.1 Uni- bi- og multivariate datasæt44

Datasæt er sæt af en eller flere variable:

- Univariate datasæt fx tider ved marathonløb
- Bivariate datasæt fx tider ved marathonløb og køn
- Multivariate datasæt fx tider ved marathonløb, køn, alder, medlem af sports klub.

### 8.0.2 Kvalitative variable33

Kvalitative variable er data vi ikke kan måle eller tælle. De antager værdier i form af navne eller labels:

- Kæledyr: kat, hund, marsvin
- Køn: mand, kvinde
- Favorit app: Angry Birds, Messenger, Audible, Tinder

### 8.0.3 Kvantitative variable33

Kvantitative variable er målbare numeriske variable, vi deler disse op i *kontinuerte* og *diskrete* variable

#### 8.0.3.1 Diskrete variable33

Diskrete variable er fx.

- Antal biler der passerer en bro observeret over flere dage.
- Dagsproduktionen af chokoladefrøer på Toms.
- Antal personer der har iPhones
- Antallet af indbyggere i en by

#### 8.0.3.2 Kontinuerte variable33

Kontinuerte variable er fx.

- Antal ml. indhold i shampoo flasker
- Aktiekurser for Intel
- Vægten på værnepligtige
- Højden på studerende

```
library(shiny)

textInput("name", "What is your name?")
```

What is your name?

```
numericInput("age", "How old are you?", NA, min = 0, max = 150)
```

How old are you?

### 8.0.4 Skaltyper<sup>33</sup>

Vi kan ydermere inddele variable efter skalatype hvor lavere betyder mindst restriktiv.

1. Nominalskala, bruges til at måle kvalitative data (er der kun 2 mulige udfald kaldes variabelen specielt binær eller dikotom), fx.
  - Køn Mand Kvinde
  - Styresystem: IOS Android Windows Symbian Andet
  - Race: Europæisk, Afrikansk, Asiatisk Andet
2. Ordinalskala inddeler data efter en rangordning
  - Karakterer på 7 trins skalaen -3 00 02...
  - Moodys credit ratings Aaa Aa A Baa Ba B Caa Ca C
  - Tilfredshed meget utilfreds, noget utilfreds, nogenlunde tilfreds, meget tilfreds
3. Intervalskala man kan sammenligne afstande og forskelle, men der er intet meningsfuldt nulpunkt. Nul for en intervalskala variabel betyder således ikke fravær af den målte størrelse. Nul grader celsius betyder altså ikke fravær af temperatur (det absolutte nulpunkt 0 Kelvin, hvor alle molekyler og atomer er i grundtilstanden). En IQ på 0 betyder ikke fravær af intelligens.
  - Temperatur målt i Celsius
  - Temperatur målt i Fahrenheit
  - PH
  - IQ
4. Ratioskala
  - Beløb i lommen
  - Højde på studerende
  - Hastighed af biler ved vejkryds
  - Indhold i Coca Cola flasker

“Statistics are used much like a drunk uses a lamppost: for support, not illumination.”

- Vin Scully

Interval- og ratioskalaer omtales som numeriske eller kontinuerte skalaer, disse er knyttet til kvantitative variable.

Nominal- og ordinalskalaer omtales ofte som kategorisk eller faktor, disse er knyttet til kvalitative variable.

En stikprøve af skalatype ratio kan fx. reduceres til ordinal, eller nominal. Temperatur målt i celsius kan fx. omskrives til en ordinal variabel: koldt normalt varmt, eller en nominal variabel: ekstrem temperatur eller normal temperatur.

Kategoriske skalaer kan yderligere reduceres til en dikotom skala, ved at sammenlægge kategorierne, til man kun har 2 kategorier.

Det er vanskeligere at ændre en nominal- ordinal- eller ratioskala til en intervalskala. At ændre variabelen nominalskala variabelen køn til ordinal giver fx. ikke mening.

#### 8.0.4.0.1 Deskriptorer Freestat 33



### 8.0.4.1 Freestat 33

Man kan få beregnet deskriptorerne i et utal af programmer heriblandt Freestat, der kan hentes herunder. Freestat kan gennemføre de mest almindelige statistiske analyser, der er flere forskellige versioner.

Hent Freestat til finansøkonomer her

Hent Freestat fuld version her

Der findes opgaver quizzes og yderligere resourcer på [www.edutest.dk](http://www.edutest.dk)

Min gode ven Benjamin Tejlbjerg har lavet en super hjemmeside med gymnasie matematik og statistik <http://www.mathhx.dk>. Siden er gratis og god til at genopfriske basisbegreber indenfor statistik, vi kommer ikke i dybden med disse begreber her.

For at få et overblik over hypoteserne for de forskellige tests kan man hente et hypotese mindmap, der er mindmaps til flere forskellige uddannelser under materialer. Her er hypoteseoversigten for finansøkonomer: CPH Finansøkonom statistik mindmap

Dette er 1. udgave af bogen, der tages forbehold for tryk og tastefejl, men alle fejl eller uklarheder i måtte finde rettes med fluks. Forslag til forbedringer modtages med kys hånd.

Der er en samlet oversigt over materialerne, der bruges i denne bog, disse ligger i som linket materialer i menuen for oven.

## 8.0.5 Hvad bruger man statistik til.33

I alle brancher i den finansielle sektor spiller statistik en rolle.

Bankerne sammensætter investeringsporteføljer, der minimerer risikoen (variansen), ved aktiver der har lav eller negativ samvariation (kovarians). Cykliske aktier som FL Smidth har fx. lav samvariation med en ikke cyklisk aktie som Novo.

Forsikringsselskaberne beregner præmier for forsikringstageren, baseret på statistiske sandsynligheder for at en hændelse indtræffer. Modellerne kan være meget specifikke, en indboforsikring kan fx. være baseret på ikke bare postnummer, boligform, men også etage, uddannelsesniveau etc.

Finansielle virksomheder underlagt finanstilsynet, bruger modeller til beregning af risiko baseret på statistisk analyse.

Mægleren beregner udbudspriser, udfra en multipel lineær regressionsmodel, der indeholder variable som størrelse, energimærke, tagtype etc.

---

*Noterne er kun til personligt brug. Alle rettigheder forbeholdes. Fotografisk eller anden gengivelse af eller kopiering eller anden udnyttelse, er uden forfatterens skriftlige samtykke forbudt ifølge dansk lov om ophavsret.*



# Chapter 9

## Datasæt og data

### 9.0.1 Uni- bi- og multivariate datasæt

Datasæt er sæt af en eller flere variable:

- Univariate datasæt fx tider ved marathonløb
- Bivariate datasæt fx tider ved marathonløb og køn
- Multivariate datasæt fx tider ved marathonløb, køn, alder, medlem af sports klub

### 9.0.2 Kvalitative variable

Kvalitative variable er data vi ikke kan måle eller tælle. De antager værdier i form af navne eller labels:

- Kæledyr: kat, hund, marsvin
- Køn: mand, kvinde
- Favorit app: Angry Birds, Messenger, Audible, Tinder

### 9.0.3 Kvantitative variable

Kvantitative variable er målbare numeriske variable, vi deler disse op i *kontinuerte* og *diskrete* variable

#### 9.0.3.1 Diskrete variable

Diskrete variable er fx.

- Antal biler der passerer en bro observeret over flere dage.
- Dagsproduktionen af chokoladefrøer på Toms.
- Antal personer der har iPhones
- Antallet af indbyggere i en by

#### 9.0.3.2 Kontinuerte variable

Kontinuerte variable er fx.

- Antal ml. indhold i shampoo flasker
- Aktiekurser for Intel
- Vægten på værnepligtige
- Højden på studerende

### 9.0.4 Skalatyper<sup>55</sup>

Vi kan ydermere inddele variable efter skalatype hvor lavere betyder mindst restriktiv.

1. Nominalskala, bruges til at måle kvalitative data (er der kun 2 mulige udfald kaldes variabelen specielt binær eller dikotom), fx.
  - Køn Mand Kvinde
  - Styresystem: IOS Android Windows Symbian Andet
  - Race: Europæisk, Afrikansk, Asiatisk Andet
2. Ordinalskala inddeler data efter en rangordning
  - Karakterer på 7 trins skalaen -3 00 02...
  - Moodys credit ratings Aaa Aa A Baa Ba B Caa Ca C
  - Tilfredshed meget utilfreds, noget utilfreds, nogenlunde tilfreds, meget tilfreds
3. Intervalskala man kan sammenligne afstande og forskelle, men der er intet meningsfuldt nulpunkt. Nul for en intervalskala variabel betyder således ikke fravær af den målte størrelse. Nul grader celsius betyder altså ikke fravær af temperatur (det absolutte nulpunkt 0 Kelvin, hvor alle molekyler og atomer er i grundtilstanden). En IQ på 0 betyder ikke fravær af intelligens.
  - Temperatur målt i Celsius
  - Temperatur målt i Fahrenheit
  - PH
  - IQ
4. Ratioskala
  - Beløb i lommen
  - Højde på studerende
  - Hastighed af biler ved vejkryds
  - Indhold i Coca Cola flasker

“Statistics are used much like a drunk uses a lamppost: for support, not illumination.”

- Vin Scully

Interval- og ratioskalaer omtales som numeriske eller kontinuerte skalaer, disse er knyttet til kvantitative variable.

Nominal- og ordinalskalaer omtales ofte som kategorisk eller faktor, disse er knyttet til kvalitative variable.

En stikprøve af skalatype ratio kan fx. reduceres til ordinal, eller nominal. Temperatur målt i celsius kan fx. omskrives til en ordinal variabel: koldt normalt varmt, eller en nominal variabel: ekstrem temperatur eller normal temperatur.

Kategoriske skalaer kan yderligere reduceres til en dikotom skala, ved at sammenlægge kategorierne, til man kun har 2 kategorier.

Det er vanskeligere at ændre en nominal- ordinal- eller ratioskala til en intervalskala. At ændre variabelen nominalskala variabelen køn til ordinal giver fx. ikke mening.

#### 9.0.4.0.1 Deskriptorer Freestat 55

Vis indhold ææø

Hello world  $\alpha$

Her skriver vi noget bla bla bla

---

```
##      speed      dist
## Min.   : 4.0    Min.   : 2.00
## 1st Qu.:12.0    1st Qu.: 26.00
## Median :15.0    Median : 36.00
## Mean   :15.4    Mean    : 42.98
```

```
## 3rd Qu.:19.0  3rd Qu.: 56.00
## Max.      :25.0  Max.      :120.00
```



# Chapter 10

## Indledning Start

Open me!

Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Vestibulum tortor quam, feugiat vitae, ultricies eget, tempor sit amet, ante. Donec eu libero sit amet quam egestas semper. Aenean ultricies mi vitae est. Mauris placerat eleifend leo. Quisque sit amet est et sapien ullamcorper pharetra. Vestibulum erat wisi, condimentum sed, commodo vitae, ornare sit amet, wisi. Aenean fermentum, elit eget tincidunt condimentum, eros ipsum rutrum orci, sagittis tempus lacus enim ac dui. Donec non enim in turpis pulvinar facilisis. Ut felis.

Spørsmål 1 whatever

Svar 1 whatever

Spørsmål whatever

Spørsmål whatever

### 10.1 Does Ruby work with knitr?

```
x = 'hello, ruby world!'
p x.split(' ')

## ["hello,", "ruby", "world!"]
```

### 10.2 js

### 10.3 Bash

```
echo "Hello Bash!"
ls

## Hello Bash!
## 01-ny_litteratur-kopi.Rmd
## 02-literature.Rmd
## 03-method.Rmd
## 04-application.Rmd
## 05-summary.Rmd
```

```
## 06-references.Rmd
## 07-application-kopi.Rmd
## DESCRIPTION
## LICENSE
## README.md
## _book
## _bookdown.yml
## _bookdown_files
## _build.sh
## _deploy.sh
## _output.yml
## book.bib
## bookdown-demo.Rmd
## bookdown-demo_files
## favicon.ico
## img
## index.Rmd
## intro.Rmd
## packages.bib
## preamble.tex
## stat2.Rproj
## style.css
## tepedulogo.svg
## toc.css
```

```
void square(double *x) {
  *x = *x * *x;
}
```

Test the `square()` function:

```
.C('square', 9)
```

```
## [[1]]
## [1] 81
```

```
.C('square', 123)
```

```
## [[1]]
## [1] 15129
```

## 10.4 A normal R code chunk

```
library(reticulate)
x = 42
print(x)
```

```
## [1] 42
```

## 10.5 Modify an R variable

In the following chunk, the value of `x` on the right hand side is 42, which was defined in the previous chunk.



```
x = x + 12  
print(x)
```

```
## [1] 54
```

## 10.6 A Python chunk

This works fine and as expected.

```
x = 42 * 2  
print(x)
```

```
## 84
```

The value of `x` in the Python session is 84. It is not the same `x` as the one in R.

## 10.7 Modify a Python variable

```
x = x + 18  
print(x)
```

```
## 102
```

Retrieve the value of `x` from the Python session again:

```
py$x
```

```
## [1] 102
```

Assign to a variable in the Python session from R:

```
py$y = 1:5
```

See the value of `y` in the Python session:

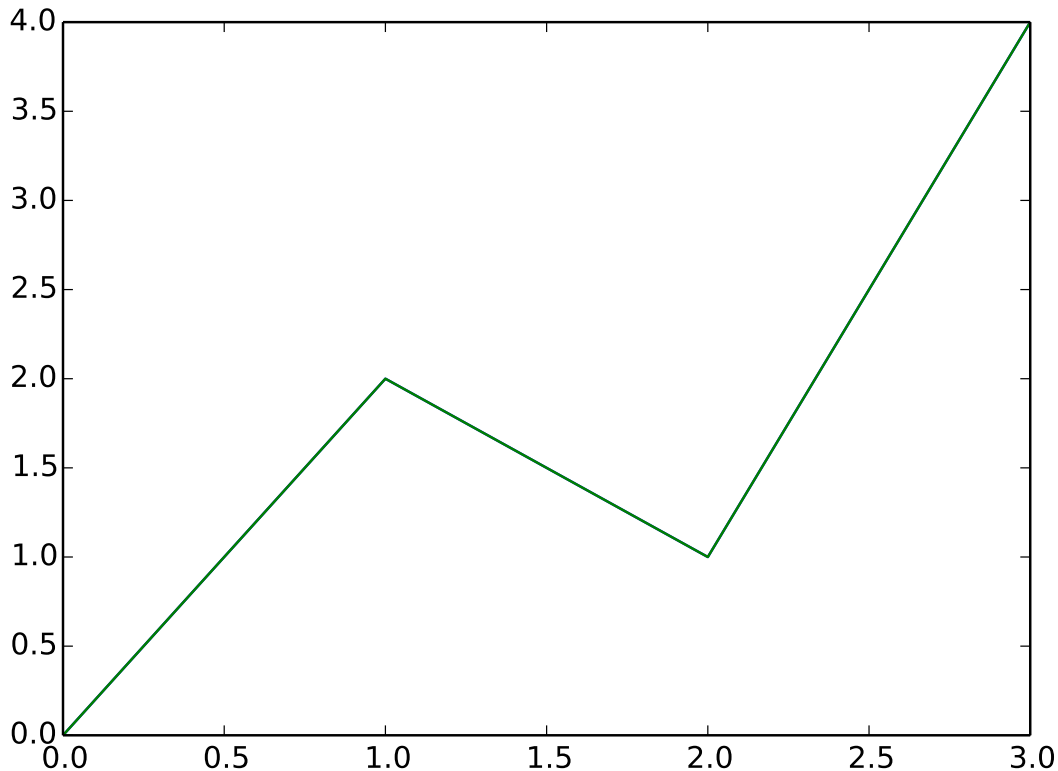
```
print(y)
```

```
## [1, 2, 3, 4, 5]
```

## 10.8 Python graphics

You can draw plots using the **matplotlib** package in Python.

```
import matplotlib.pyplot as plt  
plt.plot([0, 2, 1, 4])  
plt.show()
```



Here is an inline note.<sup>1</sup>

Here is an inline note.<sup>2</sup>

The quick brown fox<sup>3</sup> jumped over the lazy dog<sup>4</sup>.

Dette er en `<span class="note" desc="lorem ut pretium malesuada, diam nunc interdum tellus, id venenatis tortor nulla eget lorem. Aenean pharetra imperdiet aliquam. Proin luctus nibh nec lobortis mattis. Sed egestas erat et urna egestas faucibus. Fusce porttitor dapibus tincidunt. Integer congue mattis nulla iaculis mollis.`

Dette er en `<span class="note" desc="lorem ut pretium malesuada, diam nunc interdum tellus, id venenatis tortor nulla eget lorem. Aenean pharetra imperdiet aliquam. Proin luctus nibh nec lobortis mattis. Sed egestas erat et urna egestas faucibus. Fusce porttitor dapibus tincidunt. Integer congue mattis nulla iaculis mollis.`

This is some text with a marginal note. Lorem ipsum dolor sit amet, consectetur adipiscing elit. Nam consectetur, lorem ut pretium malesuada, diam nunc interdum tellus, id venenatis tortor nulla eget lorem. Aenean pharetra imperdiet aliquam. Proin luctus nibh nec lobortis mattis. Sed egestas erat et urna egestas faucibus. Fusce porttitor dapibus tincidunt. Integer congue mattis nulla iaculis mollis.

This is some text with a marginal note. Lorem ipsum dolor sit amet, consectetur adipiscing elit. Nam consectetur, lorem ut pretium malesuada, diam nunc interdum tellus, id venenatis tortor nulla eget lorem. Aenean pharetra imperdiet aliquam. Proin luctus nibh nec lobortis mattis. Sed egestas erat et urna egestas faucibus. Fusce porttitor dapibus tincidunt. Integer congue mattis nulla iaculis mollis.

<sup>1</sup>Inlines notes are easier to write, since you don't have to pick an identifier and move down to type the note.

<sup>2</sup>Inlines notes are easier to write, since you don't have to pick an identifier and move down to type the note.

<sup>3</sup>Foxes are red

<sup>4</sup>Dogs are usually not red

## 10.9 Opgaver

### 10.9.1 Spørgsmål 1

(tab content)

### 10.9.2 Svar 1

(tab content)

### 10.9.3 Spørgsmål 2

Spørgsmål 2

### 10.9.4 Spørgsmål 2

(Spørgsmål 2)

## 10.10 Quarterly Results

### 10.10.1 Spørgsmål 1

(tab content)

### 10.10.2 Svar 1

(tab content)

### 10.10.3 Spørgsmål 2

Spørgsmål 2

### 10.10.4 Spørgsmål 3

$$\frac{\beta}{\alpha + 33}$$

# End tabs

## 10.11 Nyt afsnit efter tabbed

Plain text

End a line with two spaces to start a new paragraph.

*italics* and *italics*

**bold** and **bold**

superscript<sup>2</sup>

~~strikethrough~~

link  
 # Header 1  
 ## Header 2  
 ### Header 3 #### Header 4 ##### Header 5 ##### Header 6 endash: –  
 emdash: —  
 ellipsis: ...  
 inline equation:  $A = \pi * r^2$   
 ...  
 horizontal rule (or slide break): \*\* > *block quote*  
 unordered list  
 \* item 2  
 + sub-item 1  
 + sub-item 2  
 1. ordered list  
 2. item 2  
 • sub-item 1  
 • sub-item 2

Table Header	Second Header
Table Cell	Cell 2
Cell 3	Cell 4

You can label chapter and section titles using `{#label}` after them, e.g., we can reference Chapter 7. If you do not manually label them, there will be automatic labels anyway, e.g., Chapter 3.

Figures and tables with captions will be placed in **figure** and **table** environments, respectively.

```
par(mar = c(4, 4, .1, .1))
plot(pressure, type = 'b', pch = 19)
```

Reference a figure by its code chunk label with the **fig:** prefix, e.g., see Figure 10.1. Similarly, you can reference tables generated from `knitr::kable()`, e.g., see Table 10.2.

```
knitr::kable(
  head(iris, 20), caption = 'Here is a nice table!',
  booktabs = TRUE
)
```

You can write citations, too. For example, we are using the **bookdown** package (Xie, 2018) in this sample book, which was built on top of R Markdown and **knitr** (Xie, 2015).

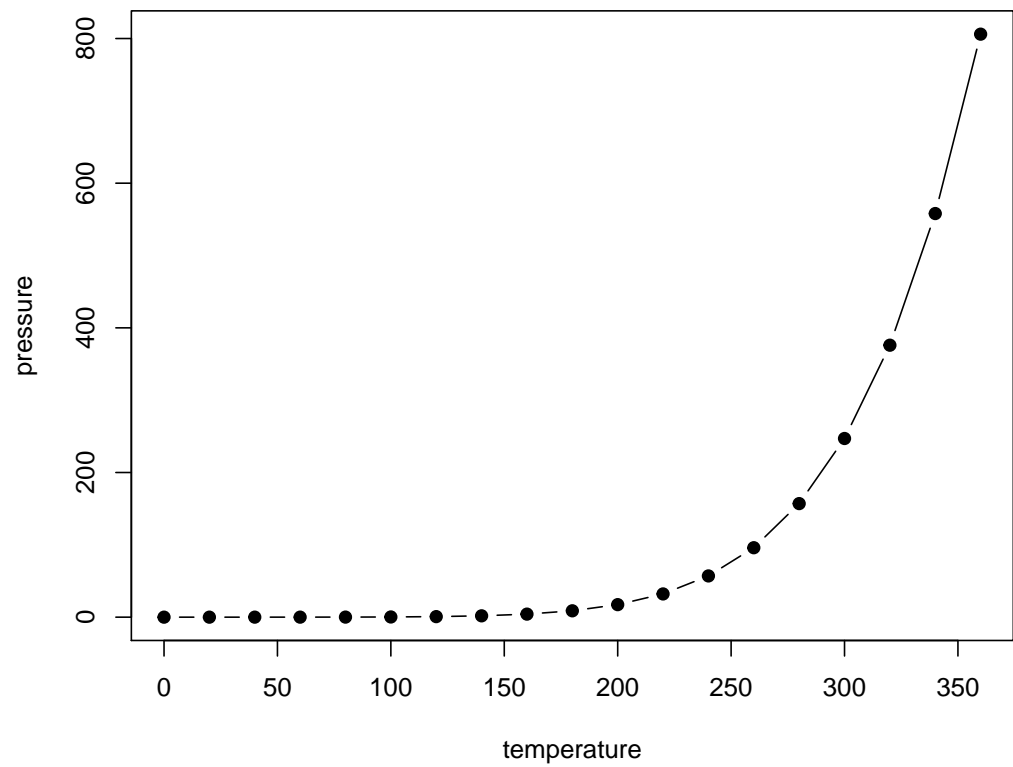


Figure 10.1: Here is a nice figure!

Table 10.2: Here is a nice table!

Sepal.Length	Sepal.Width	Petal.Length	Petal.Width	Species
5.1	3.5	1.4	0.2	setosa
4.9	3.0	1.4	0.2	setosa
4.7	3.2	1.3	0.2	setosa
4.6	3.1	1.5	0.2	setosa
5.0	3.6	1.4	0.2	setosa
5.4	3.9	1.7	0.4	setosa
4.6	3.4	1.4	0.3	setosa
5.0	3.4	1.5	0.2	setosa
4.4	2.9	1.4	0.2	setosa
4.9	3.1	1.5	0.1	setosa
5.4	3.7	1.5	0.2	setosa
4.8	3.4	1.6	0.2	setosa
4.8	3.0	1.4	0.1	setosa
4.3	3.0	1.1	0.1	setosa
5.8	4.0	1.2	0.2	setosa
5.7	4.4	1.5	0.4	setosa
5.4	3.9	1.3	0.4	setosa
5.1	3.5	1.4	0.3	setosa
5.7	3.8	1.7	0.3	setosa
5.1	3.8	1.5	0.3	setosa



# Bibliography

Xie, Y. (2015). *Dynamic Documents with R and knitr*. Chapman and Hall/CRC, Boca Raton, Florida, 2nd edition. ISBN 978-1498716963.

Xie, Y. (2018). *bookdown: Authoring Books and Technical Documents with R Markdown*. R package version 0.7.