Statistik

Thomas Petersen

2019-04-28

Contents

Adgang

.

Glemt? Vis

Login her!

Husk mig

.

Profil/Afmeld

Log ud

Javascript Required

For testing.

You must have JavaScript enabled in order to log in.

Video sådan køber du adgang.

Video sådan logger du ind.

Use `gitbook` to convert the `text` in markdown syntax to HTML.

Chapter 1

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 Bordeux 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 statistike 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.

1.1 Freestat basisversion

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

1.2 Freestat fuld version

Har du købt adgang til premium abbonnementet, er der en del ekstra analyser, derfor bør du hente Freestat premium versionen. Seneste version af programmet kan hentes her.

Du kan finde flere resourcer bagerst i bogen under materialer ved at klikke her.

Der findes opgaver quizzes og yderligere resourcer på 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.

Denne online bog rettes og opdateres løbende med nye videoer opgaver og quizzes, der tages forbehold for tryk og tastefejl, men alle fejl eller uklarheder I måtte finde rettes med fluks. Forslag til forbedringer modtages med kyshånd.

Chapter 2

Datasæt og data

Sentry Page Protection

Please Wait...

2.0.1 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

2.0.2 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

2.0.3 Kvantitative variable

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

2.0.3.1 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

2.0.3.2 Kontinuerte variable

Kontinuerte variable er fx.

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

2.0.4 Skalatyper

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 variablen 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 variablen nominalskala variablen køn til ordinal giver fx. ikke mening.

Chapter 3

Materialer

MINDMAPS og Freestat

Hent Freestat premium her
Hent Freestat basis her
CPH Finansøkonom statistik mindmap

Skabeloner

Skabelon test af middelværdi
Skabelon test af standardafvigelse
Skabelon test af andel p
Skabelon test af 2 andele
Skabeloner test 2 middelværdier
Skabelon Simpel lineær regression
Skabelon Multipel lineær regression
Skabelon Goodness of fit test
Skabelon Chi i anden test
Skabelon ANOVA/ANAVA

Datasæt

2015-Januar-Data.xlsx
AFFAIRS.xls data vedr. utroskab
BANKDATA.xls data vedr. bankansattes uddannelse, køn og etnicitet
Debt.xlsx
Forbes 400 2014 RICH US
Forbes-Global-2000 YEAR 2015
FORD FOCUS
Fortune-Global-500
GDP
GDP PER CAPITA USD
GDP2015
HELBRED
Hjemmesidedesigns besøgstider i millisekunder

Datasæt
IMDB stikprøve på 759 film
KØNSROLLER
MEDIEFORBRUG
SOMMERHUSE LEJE RØMØ
Statkarakterer
TITANIC
TYVERI
usarrest
VIRKSOMHEDER-DK
Yahooinvestexcel
USA afkast pr md
Hjemmeopgave USA afkast
Hjemmeopgave USA afkast løsning

Finansøkonom valgfag statistik eksamensopgaver
2017 December eksamensopgave valgfag statistik
2017 December data valgfag statistik
2017 November eksamensopgave valgfag statistik
2017 November data valgfag statistik
2017 Juni eksamensopgave valgfag statistik
2017 Juni data valgfag statistik
2017 Maj eksamensopgave valgfag statistik
2017 Maj data valgfag statistik
2017 Maj løsningsforslag valgfag statistik

Finansøkonom statistik eksamensopgaver
Mappe med gamle eksamensopgaver

Edutest
www.edutest.dk

Hjemmeside af Benjamin Tejlbjerg med gymnasie matematik og statistik
http://www.mathhx.dk

Chapter 4

Chi i anden tests

4.1 Goodness of fit test

Goodnees of fit testen er en udvidelse af z-testet for en andel. Med test af andele kan man fx. undersøge om andelen af mænd er 60% og kvinder 40% i en population, vi tester altså fordelingen for en kvalitativ variabel med 2 mulige udfald. Med et goodness of fit test kan vi teste kvalitative variable med 2 eller flere mulige udfald, man kan fx. undersøge om fordelingen af boligform i en stikprøve kan antages at svare til fordelingen på regionsplan: 50% ejer, 20% andel og 30% leje.

Vi tester vha. Chi i anden fordelingen. Teststørrelsen vi finder, udtrykker forskellen mellem det vi observerer i stikprøven og det vi tester under nulhypotesen.

Antag man simpelt tilfældigt har udtaget en stikprøve på 150 boliger, der indeholder 60 ejer- 40 andels- og 50 lejeboliger.

Hvis vi vil undersøge undersøge om fordelingen af boligform i stikprøven, kan antages at følge regionsfordelingen som er 50% ejer, 20% andel og 30% leje, opstiller vi følgende hypoteser:

$$H_0 : p_{ejer} = 0.5 \; p_{andel} = 0.2 \; p_{leje} = 0.3$$

$$H_1 : Fordelingen\ af\ boliger\ f\ lger\ ikke\ samme\ fordeling\ som\ i\ regionen$$

Teststørrelsen findes som:

$$\chi^2 = \sum_{j=1}^{k} \frac{(O-E)^2}{E}$$

Hvor O er observerede værdier og E er forventede værdier det stammer fra expected på engelsk, k angiver antallet af mulige udfald for den kvalitative variabel.

For at beregne teststørrelsen bestemmer vi E, antallet af ejer, leje og andel vi ville forvente i en stikprøve på netop 150 boliger, der perfekt repræsenterede regionen.

ejer: $0.5 \cdot 150 = 75$ andel: $0.2 \cdot 150 = 30$ leje: $0.3 \cdot 150 = 45$

Vi kan nu udregne teststørrelsen som: