# Contents



# 1 Disclaimer

Disse noter blev udarbejdet i forbindelse med jeg underviste i kurset **Sandsynlighedsteori og statistik** udbudt af Økonomisk Institut, Københavns Universitet.

Dette er ikke blevet gennemlæst, rettet eller på anden måde redigeret af en tredje person, som ville kunne fange evt. fejl og mangler. Derfor **forvent** at der er fejl i dette dokument. Forhold dig kritisk til resultaterne, og hvis du er sikker på der er en fejl, så tag udgangspunkt i det.

Dokumentet indeholder rettevejledninger til øvelsesseddlerne forbundet med faget. Der er et tilhørende github-repository:

*https://github.com/JakartaLaw/statistik2018*.

# Lecture Notes

Jeppe Johansen

November 6, 2018

# 2 Topics

## 2.1 Deskriptiv statistik

Vi har indsamlet noget data:

$$y_1, y_2, \cdots, y_n \tag{1}$$

Man kan forestille sig en DGP (data generende proces) have forskellige karakteristika, hvormed den mapper til en respons variabel, som kan være:

- binær
- tælle
- diskret
- kontinuær

frekvens for $j$'te element i $\mathbb{Y}$ udregnes ved:

$$f_{y=j} \sum_{i=0}^{N} \mathbb{1}(y_i = j) \tag{2}$$

Den empiriske cumulative distribution

$$F(y) = \sum_{i=1}^{N} \frac{\mathbb{1}(y_i \leq y)}{n} \tag{3}$$

Empriske momenter:

- mean (gennemsnit) $= \frac{1}{N}\sum_{i=1}^{N} y_i$
- Varians $= \frac{1}{N}\sum(y_i - \bar{y})^2$
- Skewness $= \frac{1}{N}\sum(y_i - \bar{y})^3$
- Kurtosis $= \frac{1}{N}\sum(y_i - \bar{y})^4$

standard afvigelse:

$$std(Y) = \sqrt{Var(Y)} \tag{4}$$

Excess Kurtosis vil sige Kurtosis - 3. Da en standard normal fordeling har kurtosis på 3.

Man vil ofte standardisere data når man kigger på skewness og kurtosis.

Standardisering af data er:

$$y_{i,standardised} = \frac{y_i - \bar{y}}{std(y)} \tag{5}$$

hvor *std* betyder standard deviation (standard afvigelse).

# 3 Lectures

---

## 3.1 Øvelse 17

#### 22/10/2018, opgaver: 1, 2, 3

### 3.1.1 Opgave 1

Kig do file do_1_17

### 3.1.2 Opgave 2

**Del 1**

Figur A: Histogram (og kernel density estimation). Viser tæthedsfunktionen (eller en approksimation).

Figur B: Viser den empiriske CDF.

Figur C: Q-Q plot er et plot der viser der modholder den empiriske distribution med en parametrisk - i dette tilfælde den gaussiske distribution. Dette er gjort ved at sammenligner quantiler.

Figur D: Boxplot - giver indblik i antal outliers samt hvordan de kvartiler, og median er fordelt.

**Del 2**

A) Medianen er den observation som er svarer til det punkt hvor $F(x) = 0.5$. Altså med andre ord $F^{-1}(0.5) = \text{median}$

B) Kig boxplot. Ja det synes der at være. Det er dog altid svært at vurdere outliers.

C) Ja, vi ser at fordelingen er centreret omkring en middelværdi og er stort set symmetrisk og unimodal.

D) Kig CDF. Omkring halvdelen af alle firmaerne.

C) 10% fraktilen (hvilket kaldes 10% percentilen). Angiver det punkt hvor $F(x) = 0.1$. I vores konkrete tilfælde cirka $-500$

E) Vi kan se på Q-Q plottet at distribution er lidt lang i halerne, men det er ikke klart om den er venstre eller højre skæv. Derudover er det ikke klart om disse afvigelser i halerne er nok, til at antage den skulle være venstre skæv eller højre skæv.

### 3.1.3 Opgave 3

**Del 1**

*gns_gym*: Man finder at gennemsnittet fra gymnasiet er kontinuær.

*studietimer*: Tælle data.

Man finder at genn