

## Anvendelse af DataMeter i et statistikforløb om konditionstal

Disse sider fungerer som en overordnet beskrivelse af et forløb med anvendelse af DataMeter i forbindelse med ugrupperede data – her konditionstal. Undervejs i siderne vil der blive henvist til et DataMeter-ark, som skal udleveres/bruges i løbet af forløbet. Det antages, at der er computere og netadgang til rådighed under hele forløbet.

### **Litteratur:**

- (1) Statistik i Løb, af Blæsild og Kristensen, LMFK 2006
- (2) MATA2, af Carstensen, Frandsen og Studsgaard, Systime, 2006

Forløbet består af *3 dele*, og i alt har et omfang på cirka 9 lektioner a 90 min.

**Del 1:** Afklaring af, hvad statistik er og hvad det kan!

**Del 2:** Beskrivende statistik – præsenter data!

**Del 3:** statistisk analyse – teser og konklusioner.

### **Om del 1 (1 lektion):**

Den første del af forløbet tager udgangspunkt i en afsøgning af emnet. Det antages her, at klassen indledningsvist allerede har arbejdet med sandsynlighedsregning og de begreber der hører til her (stokastisk variabel, etc.)

Der arbejdes i grupper ud fra følgende (arbejds)spørgsmål.

**1a)** Hvad er statistik – og hvordan kan den bruges i forhold til virkeligheden? (brug gerne Internettet til at finde informationer)

**1b)** Hvis I bliver bedt om at lave en opsummerende beskrivelse – også kaldet en *beskrivende statistik* – hvad er så vigtigt? Hvad vil I tage med? Og hvad giver et godt overblik over og god information om data?

Delen slttes af med en kort diskussion i plenum af de to dele – deskriptiv statistik og statistisk analyse.

### **Om del 2 (3 lektioner):**

En vigtig del af statistikken er den grafiske del. Denne del kræver en introduktion til DataMeter's grafiske evner, og dette illustreres indledningsvist af læreren på *andre data*, end dem eleverne får udleveret.

Desuden diskuteres og præsenteres den *teoretiske baggrund* bag fraktiler og frekvens ud fra (1) og (2) i denne fase – og dermed medianer og kvartiler. Dette kan gøres ud fra opgaver og udregninger.

Delen sluttet af med, at eleverne får udleveret DataMeter-arket "kondital", som de grundet den indledende fase nu kan bruge til at præsentere data og lave indledende konklusioner på.

Mulige arbejds(spørgsmål):

**2a)** Lav en fempunktsopsummering for både drenge og piger. Lav og kommenter på boxdiagrammerne.

**2b)** Lav sumkurven og udregn forskellige fraktiler og frekvenser. Kommenter på disse.

Delen rundes af med en diskussion af, hvad "der mangler" for at statistikken "har gjort en forskel"...

### **Om del 3 (5 lektioner):**

Indledningsvist præsenteres teorien bag de *empiriske størrelser* fra (1) og (2) og disse udregnes vha DataMeter (hvis dette ikke allerede er sket) for det aktuelle datasæt (kondital).

Eleverne har desuden inden denne del startes op læst om *normalfordelingen* i (1) og (2) og arbejder derefter med følgende spørgsmål:

**3a)** Hvad er hele ideen med at introducere normalfordelingen?

**3b)** Hvad er sammenhængen mellem tæthedsfunktion  $f(x)$  og fordelingsfunktionen  $F(x)$ ?

**3c)** Hvilken sammenhæng er der mellem standardnormalfordelingen og alle andre normalfordelinger?

Herefter tegnes der fraktildiagrammer og anvendeligheden af disse – og modelbegrebet - gennemgås i plenum på baggrund af (1) og ovenstående spørgsmål. DataMeter-filen "kondital med tæthedsfunktion drenge" inddrages i denne diskussion.

Eleverne får nu lidt tid til at "summe" over følgende:

**3d)** Hvilke (*hypo*)teser kunne det være interessante at undersøge (teste) ud fra de datasæt I har fået udleveret? Hvordan vil I gøre det? Hvilke fejlkilder vil I tænke over?

Herefter gennemgås *t-test teorien* og *konfidensinterval* ud fra (1). Der arbejdes nu med følgende spørgsmål (DataMeter inddrages, hvor det findes passende).

**3e)** Betragt drengenes kondital. Vi har vist ud fra fraktildiagrammerne, at disse data er normalfordelte, altså  $M_0: X_i \sim N(\mu, \sigma^2)$ . I denne model vil vi gerne

teste hypotesen  $H: \mu=46$ . Hvad betyder denne hypotese? Og hvad betyder det, hvis vi forkaster hypotesen i et test?

**3f)** Udregn t-teststørrelsen for testet fra 3e). Regn altså  $t(x_1, x_2, \dots, x_{74})$ . Forsøg at forklare hvad den udregnede værdi siger noget om. Hvordan skal den forstås?

**3g)** Forkaster eller accepterer I hypotesen på baggrund af testet i 3f)? Brug skemaet bag i (1) eller lommeregnerens/DataMeters "Inverse t" funktion til at vurdere dette. Hvilken model har vi nu?

**3h)** Det skulle gerne være klart, at testet i 3f) ikke er helt *fyldestgørende* for hvilken information vi gerne vil have ud af observationerne. Derfor introduceres 95 % konfidensintervallerne. Udregn disse i DataMeter for begge køn og forklar hvilken information de indeholder.

---

Det opfordres til, at forløbet afsluttes med en rapport, der desuden indeholder en diskussion af stikprøver – gerne på baggrund af følgende udsagn:

"Alle gymnasieelever på Egå Gymnasium er i god form"

"Gymnasieelever er i god form"

"Piger under 20 år er i dårlig form"

"Drenge under 20 år er i god form"

/Lars Bo Kristensen, juni 2008