

# Statistik og Databehandling på en TI-83 (2008 version)

Af Jonas L. Jensen (jonas@imf.au.dk).

## 1 Fordelingsfunktioner

Husk på, at en fordelingsfunktion for en stokastisk variabel  $X$  er funktionen

$$F_X(t) = P(X \leq t)$$

og at variable med samme fordeling har samme fordelingsfunktion. Disse kan vi finde på vores TI-83 (eller lignende) ved hjælp af cdf-kommandoerne, der findes under DISTR (2nd + VARS). Det gør vi på følgende måde

**Normalfordeling** Lad  $X \sim N(\mu, \sigma^2)$ . Så er

$$F_X(x) = \text{normalcdf}(-10^{99}, x, \mu, \sqrt{\sigma^2})$$

Hvis i vil finde  $P(a \leq X \leq b)$  gøres det på følgende måde

$$P(a \leq X \leq b) = \text{normalcdf}(a, b, \mu, \sqrt{\sigma^2}).$$

Husk at hvis  $\mu = 0$  og  $\sigma^2 = 1$  er  $F_X(x) = \Phi(x)$ . I dette tilfælde kan jeres lommeregner også finde fraktilen, idet

$$u_\alpha = \Phi^{-1}(\alpha) = \text{invNorm}(\alpha).$$

**t-fordeling** Lad  $X \sim t[f]$ . Så er

$$F_X(x) = \text{tcdf}(-10^{99}, x, f)$$

**$\chi^2$ -fordelingen** Lad  $X \sim \chi^2[f]$ . Så er

$$F_X(x) = \chi^2\text{cdf}(0, x, f)$$

**F-fordelingen** Lad  $X \sim F[r, s]$ . Så er

$$F_X(x) = \text{Fcdf}(0, x, r, s)$$

**Binomial-fordelingen** Lad  $X \sim \text{binomial}(n, p)$ . Så er

$$F_X(x) = \text{binomcdf}(n, p, x)$$

Man kan også finde sandsynlighedsfunktionen  $P(X = x) = \text{binompdf}(n, p, x)$ .

**Poission-fordelingen** Lad  $X \sim \text{poisson}(\lambda)$ . Så er

$$F_X(x) = \text{poissoncdf}(\lambda, x)$$

Man kan også finde sandsynlighedsfunktionen  $P(X = x) = \text{poissonpdf}(\lambda, x)$ .

**Fraktiler** Når i laver tests er det som oftes fraktilen i skal bruge, og det er kun når det er en normalfordeling vi kigger på, at den kan findes direkte. Heldigvis findes der en metode, der bruger TI-83'eren's Solver funktion. Tryk **MATH** og vælg **0:Solver...** Tryk nu op indtil du har overskriften **EQUATION SOLVER**. Hvis du fx vil finde en fraktil for  $t$ -fordelingen,  $t(0.65, 11)$  indtaster du således at der kommer til at stå **eqn:0=tcdf(-10^{99},x,11)-0.65**. Tryk nu ned og indtast under **X=** et tal tæt på hvad du tror fraktilen er – her kunne det være 0.4, kig evt i tabellen bag i bogen. Tryk nu på **ALPHA**-tasten og så på **ENTER**. Nu står den og tygger lidt på det, og skulle efter nogle sekunder gerne komme med resultatet  $x = 0.39555062419845$ .

Metoden kan selvfølgelig også bruges for andet end fraktiler i  $t$ -fordelingen, men metoden er den samme.

## 2 Data

På TI-83'eren har man muligheden for at gemme en masse data, for derefter at analysere det. Tryk **STAT** og derefter **Edit...** Nu kan du flytte rundt, og fylde data i dine lister. Her vil jeg henvise til manualen (12-20).

Når vi har indtastet data, er vi klar til at analysere. Tryk **STAT** og tryk en gang til højre så vi er i menuen **CALC**. Vælg nu **1:1-Var-Stats**, og indtast så hvilken liste vi havde vores data i. Hvis det fx var i  $L_1$ , så tryk **2nd+1**. Tryk nu på enter. Nu dukker der en masse op. Bl.a.  $\bar{x}$  (skønnet af middelværdien),  $\Sigma x$  (summen af data),  $\Sigma x^2$  (summen af data i anden),  $Sx$  (skøn af standardafvigelsen, denne kalder vi normalt  $s$ ), samt  $Med$  (medianen). Hvis i har to lister med observationer, kan i vælge **2:2-Var-Stats**, og så indtaste en 2 lister adskilt af , og derefter trykke enter. Så får i noget tilsvarende.

## 3 Test og konfidensintervaller

Man kan lave test både ud fra data man har i en liste, eller ud fra udregnede estimatorer. Vi vil nu følge Jens' noter "Estimation og test: katalog". Alle de statistik-kommandoer vi refererer til, ligger under **STAT**, under menuen **TESTS**. Jeg vil her angive, hvilke kommandoer på lommeregneren, der svarer hvilke test i Jens' noter.

Men først en smule info om, hvordan man gør. I alle test kan man øverst vælge mellem **DATA** og **STATS**. Forskellen er om man har data i en liste, eller om man bare kender nogle estimatorer. Derefter skal i enten indtaste jeres hypoteseværdi, samt enten data eller estimatorer. Læg mærke til, at i skal angive estimatet for spredningen (dvs.  $s$ ) og ikke estimatet for variansen ( $s^2$ ), men det kan jo klares med en lille kvadratrods. Til sidst skal i vælge jeres alternativ-hypotese.

Når i skal finde konfidensintervaller, gøres det på nogenlunde samme måde, hvor i så også skal vælge jeres  $1 - \alpha$  under **C-level**. Som regel er denne 0.95.

Jeg vil her følge appendix i Jens Ledets noter 'Et Nanokursus i Statistik' og bruge opgaver derfra som eksempler.

### A.1 Binomialfordelt data

Hvis vi har en binomialfordeling og vil teste  $p = p_0$  gøres det med **5:1-PropZTest**. Nu skal vi indtaste  $p_0$ , observationen  $x$  og antalsparameteren  $n$ , samt vælge

alternativ-hypotesen. Konfidensintervallet for  $p$  findes med **A:1-PropZInt**. Lad os tage et lille eksempel.

**Opgave 20** I 1660 i London blev der født 3724 drenge og 3247 piger, dvs 6971 i alt. Vi lader  $p$  være sandsynligheden for at føde en dreng og vil teste  $p = 0.5$ . Gå nu ind under **5:1-PropZTest**. Her indtaster vi  $p_0 = 0.5$ ,  $x = 3724$  og  $n = 6971$ . Som alternativhypotese vælger vi  $\neq p_0$ . Flyt nu markøren til **Calculate** og tryk **ENTER**. Nu får i følgende output

- Teststørrelsen:  $z = 5.713087001$
- p-værdien  $p = 1.1126817 \cdot 10^{-8}$
- Estimat for  $p$   $\hat{p} = 0.5342131688$

Ud fra den lave p-værdi kan vi afvise hypotesen. Et konfidensinterval for  $p$  kan findes ved hjælp af **A:1-PropZInt**. Indtast nu  $x = 3724$ ,  $n = 6971$  og  $C - Level = 0.95$ . Flyt markøren til **Calculate** og tryk på **ENTER**. Outputtet er konfidensintervallet  $(0.5225, 0.54592)$ .

## A.2 Sammenligning af to binomialfordelinger

Hvis i er blevet givet to binomialfordelinger og skal teste  $p_1 = p_2$ , gøres det med **6:2-PropZTest**. . . . Nogle gange er der i en opgave angivet  $\hat{p}_1$  og  $\hat{p}_2$ , men i testen skal i skrive det observerede antal  $x_1$  og  $x_2$ . Dem kan i naturligvis finde ved

$$x_i = \hat{p}_i \cdot n_i,$$

og så runde af til et helt tal. Konfidensintervallet for  $p_1 - p_2$  findes med **B:2-PropZInt**. . . .

## A.3 Et datasæt fra normalfordeling

Hvis i har indtastet jeres data i en liste, fx  $L_1$ , kan i lave fraktilsammenligning ved at trykke **2nd** og så trykke på **Y=**. Vælg nu den første og tryk enter. Sørg for at **ON** er markeret, og vælg den sidste graf-type. Indtast  $L_1$  under **Data List**, vælg  $x$ -aksen som data-akse og vælg hvad der skal markere et punkt. Tryk nu på **ZOOM** og vælg **9:ZoomStat**. Så skulle i gerne se nogle punkter, og nu skal i så vurdere om de ligger pænt omkring en linje. I kan aflæse punkterne ved at trykke på **TRACE** og så rykke til højre og venstre mellem punkterne og aflæse værdierne i bunden.

Hvis i har skal teste  $\mu = \mu_0$  skal det gøres med **2:T-Test**. . . . Nu kan du vælge at bruge enten indtastet data (**Data**) eller indtaste dine estimationer (**Stats**). Lad os tage et par eksempler

**Data (Forår 2004, opg 2)** Tast de 10 observationer ind i liste  $L_1$  som beskrevet i afsnittet "Data". Gå nu ind under **2:T-Test**. . . og vælg menuen **Data**. Nu skal vi indtaste  $\mu_0 = 0$ . Under **List** skal i vælge i hvilken liste jeres data er - i vores tilfælde  $L_1$ . **Freq** skal være 1. Under  $\mu$ : skal i vælge hvad jeres alternativ-hypotese skal være, som regel  $\neq \mu_0$ . Når alt dette er på plads skal i flytte cursoren ned til **Calculate** og trykke enter. Nu får udregnet følgende

- Teststatistikken:  $z = 3.679915895$

- p-værdien:  $p = 0.0050761326$
- Gennemsnittet:  $\bar{x} = 2.33$
- Estimat for  $\sigma$ :  $Sx = 2.002248736$

Man kan desværre ikke teste  $\sigma^2 = \sigma_0^2$  på en TI-83, men i kan stadig få hjælp til jeres udregninger, da men fx kan finde  $F_{\chi^2(f)}(fv)$  ved at skrive  $\chi^2cdf(0, f \cdot v, f)$  (beskrevet i afsnittet om fordelingsfunktioner).

#### A.4.1 Test for ens varianser

Her skal vi bruge `D:2-SampFTest`.... Læg mærke til, at i skal indtaste data eller estimater for to observationerækker.

#### A.4.2 Test for ens middelværdier når varianserne er ens

Til dette skal i bruge `3:2-SampTTest`.... Her skal i vælge `Yes` ved `Pooled`, da vi kan antage at varianserne er ens. Brug `0:2-SampTInt`... til at finde konfidensinterval for  $\mu_1 - \mu_2$ .

#### A.4.3 Test for ens middelværdier når varianserne er forskellige

Her skal i, som ovenfor, bruge `3:2-SampTTest`..., men her skal i så vælge `No` ved `Pooled`. Brug `0:2-SampTInt`... til at finde konfidensinterval for  $\mu_1 - \mu_2$ .

### A.5 Lineær Regression

Dette gøres med `E:LinRegTTest`.... Her kan vi kun teste  $\beta = 0$ , men det er som regel også det i skal teste. Læg mærke til, at jeres X-liste og Y-liste passe sammen, dvs. at første indgang i X-listen skal svare til første indgang i Y-listen osv. Se bort fra, at der står noget med  $\rho$  under alternativ-hypotesen.

#### Andre lineær regressions test

Det kan vi desværre ikke på en lommeregner. Men til udregning af p-værdierne, kan vores afsnit om fordelingsfunktioner være til nytte.

## 4 Lineær regression

Når i har en liste med x'erne og en liste med y'erne, kan i lave lineær regression på jeres lommeregner. Det ved at trykke `STAT`, vælge `CALC` og så trykke `4:LinReg(ax+b)`. Nu skal du indtaste din X-liste hhv. Y-liste adskilt af `,`.

## 5 Afsluttende bemærkninger

For god ordens skyld skal jeg vel lige sige, at bare fordi ens lommeregner kan så mange fine ting, er det ingen undskyldning for ikke at forstå stoffet. Hvis i har kommentarer, spørgsmål eller tilføjelser, kan i sende dem til [jonas@imf.au.dk](mailto:jonas@imf.au.dk).