Stastistik og Databehandling på en TI-83

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

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

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

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

t-fordeling Lad $X \sim t[f]$. Så er

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

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

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

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

$$F_X(x) = F\mathit{cdf}(0, x, r, s)$$

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

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

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

Possion-fordelingen Lad $X \sim \mathit{poisson}(\lambda)$. Så er

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

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

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), Σx (summen af data), Σx^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 estimater. Derefter skal i enten indtaste jeres hypoteseværdi, samt enten data eller estimater. 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 kvadratrod. 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 gennemgå hvordan i gør i par tilfælde.

3.1 Normalfordeling: Teste $\mu = \mu_0$, σ^2 kendt

Vælg `1:Z-Test...`. Nu kan du vælge at bruge enten indtastet data (`Data`) eller indtaste dine estimationer (`Stats`). Lad os tage et par eksempler

Data (opg 10.3) Tast nu vores 18 observationer ind i liste L_1 som beskrevet i afsnittet "Data". Gå nu ind under `1:Z-Test...` og vælg menuen `Data`. Nu skal vi indtaste μ_0 og vores kendte spredning σ (læg mærke til, at det er σ, og ikke σ^2).

Under `List` skal i vælge i hvilken liste jeres data er - i vores tilfælde L_1. `Freq` skal være 1. Under μ: skal i vælge hvad jeres alternativ-hypotese skal være, som regel $\mu \neq \mu_0$. Når alt dette er på plads skal i flytte cursoren ned til `Calculate` og trykke enter. Nu får udregnet følgende

- Teststatistikken: $z = 1.6$
- p-værdien: $p = .1095985788$
- Gennemsnittet: $\bar{x} = 25.05333333$

- Estimat for σ: $Sx = .1373274318$

I dette tilfælde er vores p-værdi større end 0.05, så vi kan acceptere vores hypotese.

Stats (opg 10.3) Dette gøres på nøjagtigt samme måde som ovenfor bortset fra, at man skal indtaste gennemsnittet ($\bar{x}$) istedet for sin liste. Så indtast som ovenfor, men skriv $\bar{x} = 25.05333333$, og vælg `Calculate`. I skulle nu gerne få samme output som ovenfor, bortset fra at den ikke estimerer spredningen.

Konfidensinterval I kan finde konfidensintervallet under `7:ZInterval...`. Med $\alpha = 0.05$ fårmed data fra opg 10.3 dette output: `(24.961,25.146)`.

3.2 Normalfordeling: Teste $\mu = \mu_0$, σ^2 ukendt

Dette gør man under `2:T-Test...`. Fremgangsmåden er nøjagtig som i afsnittet med ukendt varians, udover at man ikke skal indtaste spredningen under `Data`, og at man skal indtaste sit estimat for spredningen under `Stats`. Hvis man indtaster data fra opg 10.3 får man følgende output

- Teststatistikken: $t = 1.647698257$
- p-værdien: $p = .117772002$
- Gennemsnittet: $\bar{x} = 25.05333333$
- Estimat for σ: $Sx = .1373274318$

Konfidensinterval Her findes konfidensintervallet under `8:TInterval...`.

3.3 Normalfordeling: Teste $\sigma^2 = \sigma_0^2$

Dette kan desværre ikke gøres på en TI-83, men i kan stadig få hjælp til jeres udregninger, da men fx kan finde $F_{\chi^2[n-1]}((n-1)v)$ ved at skrive $\chi^2 cdf(0,(n-1)\cdot v, n-1)$ (beskrevet i afsnittet om fordelingsfunktioner).

3.4 Binomialfordelingen: Teste $p = p_0$

Dette gøres med `5:1-PropZTest`. Nu skal vi indtaste vores p_0, vores observation x og n, samt vælge alternativ-hypotesen.

Eksempel: opg. 10.5 Her er $p_0 = 0.008$, $x = 11$ og $n = 800$. Vores alternativhypotese er $p \neq p_0$. Tryk nu calculate, så skulle i gerne få følgende output

- Teststatistikken: $z = 1.825626826$
- p-værdien: $p = .0679063762$
- Gennemsnittet: $\hat{p} = 0.01375$
- Antal elementer $n = 800$

Konfidensinterval Konfidensintervallet findes her med `A:1-propZInt...`.

3.5 To normalfordelinger: teste $\mu_1 = \mu_2, \sigma_x^2 = \sigma_y^2$

Til dette skal i bruge `3:2-SampTTest...`. Læg mærke til, at i skal indtaste data eller estimater for to observationer. Her skal i vælge `Yes` ved `Pooled`, da vi kan antage at varianserne er ens.

Konfidensinterval Brug `0:2-SampTInt...`.

3.6 To normalfordelinger: teste $\mu_1 = \mu_2, \sigma_x^2 \neq \sigma_y^2$

Her skal i, som ovenfor, bruge `3:2-SampTTest...`, men her skal i så vælge `No` ved `Pooled`.

Konfidensinterval Brug `0:2-SampTInt...`.

3.7 To normalfordelinger: teste $\sigma_x^2 = \sigma_y^2$

Her skal vi bruge `D:2-SampFTest...`.

3.8 To binomialfordelinger: teste $p_1 = p_2$

Dette gøres med `6:2-PropZTest...`. Ofte får i 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.

Konfidensinterval Brug `B:2-PropZInt...`.

3.9 To binomialfordelinger: log odds ratio

Den kan ikke laves på en TI-83.

3.10 Lineær regression: teste hældning

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-lister og Y-lister 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 ρ under alternativ-hypotesen. Lad os lige tage opg. 14.6 som et eksempel.

Opg 14.6 Indtast husstørrelserne (x_i'erne) i L_1 og og huspriserne (y_i'erne) i L_2. Under `E:LinRegTTest...` sætter i nu `Xlist` til L_1 og `Ylist` til L_2. Alternativhypotesen skal være $\beta \neq 0$. Tryk nu `Calculate`. Så skulle i gerne få følgende output

- $t = 5.542802202$ (teststatistikken)
- $p = .0014557765$ (p-værdien)
- $df = 6$ (dette er bare $n - 2$)

- $a = -.0893521966$ (dette er $\hat{\gamma}$)
- $b = .0106478034$ (dette er $\hat{\beta}$)
- $s = .1244496227$ (dette er vores spredningsestimat s)

Da p-værdien er under 0.05, kan vi forkaste vores hypotese om $\beta = 0$, så der er altså en lineær sammenhæng mellem huspriser og husstørrelser.

3.11 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 `,`. Lad os tage opg. 14.6 som et eksempel.

Indtast huspriserne under L_2 og størrelserne under L_1. Vælg så `8:LinReg(ax+b)` som beskrevet ovenfor, og tryk `L1,L2`. Nu skulle der gerne stå `LinReg(a+bx) L1,L2`. Tryk nu enter. Nu får vi følgende output

- $a = -.0893521966$ (dette er vores $\hat{\gamma}$)
- $b = .0106478034$ (dette er vores $\hat{\beta}$)

Se i øvrigt også 3.10 om regressionstest.

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 eller tilføjelser, kan i sende dem til jonas@imf.au.dk.