Nyheder i Gym17

Contents

1 Nyheder i Gym-pakken til Maple 17 1
1.1 Statistik 1
1.2 Regressioner 4
1.3 Vektorregning 5
1.4 Desriptiv statistik 6
Index 7

1 Nyheder i Gym-pakken til Maple 17

1.1 Statistik

Statistiske funktioner

Normalfordelingen, t-fordelingen og χ^2-fordelingen findes nu i Gym-pakken. Herefter er det ikke længere nødvendigt at skulle indlæse Statistics-pakken og definere stokastiske variabler for at kunne bruge disse fordelinger og deres inverse.

Statistiske tests

z-test - Test for middelværdi med kendt spredning

$with(Gym)$:

$obs := [229.4, 229.7, 230.2, 230.2, 232, 231.2, 230, 230.6, 230, 229.4, 230.9, 228.5, 231.5, 230.9, 231.2, 227.9, 230.6, 232, 230.3, 232.3]$:

Det vides, at observationerne er normalfordelte med med spredning $\sigma = 1.5$.

Test af hypotesen

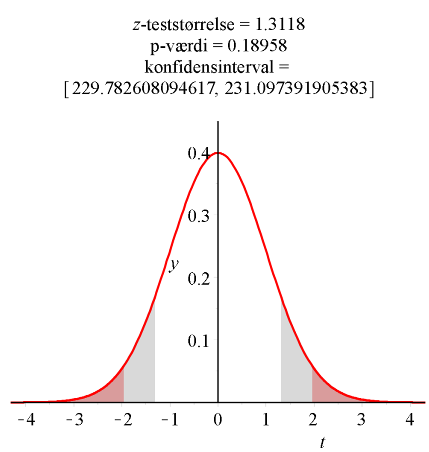
$H_0 : \mu = 230$

mod alternativet

$H_0 : \mu \neq 230$

på niveau 5%, sker med kommandoen

$zTest(obs, 230, 1.5)$



Output viser den kritiske mængde, p-værdien og 95%-konfidensintervallet.

Er der kun behov for at bestemme 95%-konfidensintervallet, sker dette med kommandoen

$zInterval(obs, 1.5, 0.95)$

$$[229.782608094715, 231.097391905285] \tag{1.1}$$

t-test - test for middelværdi med ukendt spredning

$obs := [5, 4.4, 5.7, 5.6, 5.5, 5.2, 5.0, 4.8, 3.6, 4.1, 4.6, 4.9, 4.0, 6.7, 5.5, 5.4, 6.7, 5.8, 5.4, 4.8, 5.9, 5.1, 3.8, 4.1, 6.7]:$

Det vides, at observationerne er normalfordelte. Test af hypotesen

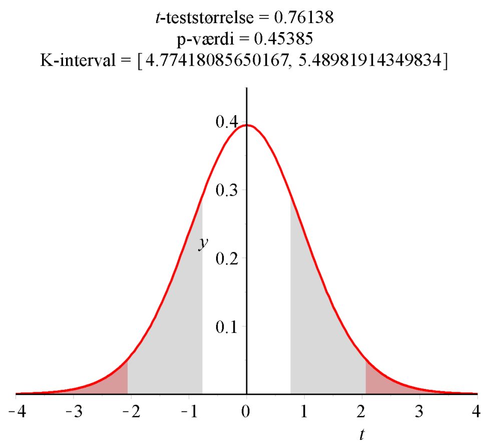
H_0: $\mu = 5$

mod alternativet

H_0 : $\mu \neq 5$

på niveau 5%, sker med kommandoen

$tTest(obs, 5)$



Er der kun behov for at bestemme fx 90%-konfidensintervallet, sker dette med kommandoen

$tInterval(obs, 0.9)$

$$[4.83538350334085, 5.42861649665915] \quad (1.2)$$

Konfidensinterval for andelen p i en binomialfordeling

Antag at vi en stikprøve på n = 180 personer har x = 40 personer der er imod højere skat. Et konfidensinterval for andelen p af personer, der er imod højere skat beregnes

$zInterval1Andel(40, 180, 0.95) = [0.161488017783562, 0.282956426616438]$

Test for lineær afhængighed

Vi tester, om der er er en lineær ahængighed mellem dataene

$X := [56, 62, 65, 80, 76, 71, 64, 75]:$

$Y := [20, 24, 28, 34, 28, 27, 23, 31]:$

$testLin(X, Y)$

	a	b
Koefficient	0.503607	-7.685012

```
 Standardfejl       0.088953       6.142175
       t-stat       5.661503      -1.251187
      p-værdi       0.001305       0.257440
  Nedre 95.00%      0.285948     -22.714330
   Øvre 95.00%      0.721266       7.344306
Frihedsgrader              6
```

Her testes hypotesen $a = 0$ (mod $a \neq 0$), og af p-værdien ses, at hypotesen $a = 0$ (altså: ingen lineær sammenhæng) må forkastes på niveau 5%. Et kig på koefficienten og konfidensintervallet viser det samme.

Ønsker vi at teste hypotesen $a = 0.3$ (mod $a \neq 0.3$), sker det således

$testLin(X, Y, a = 0.3)$

```
                           a
  Koefficient       0.503607
 Standardfejl       0.088953
       t-stat       2.288929
      p-værdi       0.062035
  Nedre 95.00%      0.285948
   Øvre 95.00%      0.721266
Frihedsgrader              6
```

Her viser p-værdien viser, at hypotesen ikke kan forkastes på niveau 5%.

Helt tilsvarende kan TestLin bruges til at teste hypoteser om b .

bidrag

- en hjælpefunktion til bestemmelse af de enkelte 'cellers' bidrag til χ^2-teststørrelsen.

$$A := \begin{bmatrix} \text{"Observeret"} & \text{"Aviser eller blade"} & \text{"Internet eller mobil"} & \text{"Radio eller TV"} & \text{"i alt"} \\ \text{"Kvinde"} & 171 & 103 & 369 & 643 \\ \text{"Mand"} & 97 & 80 & 180 & 357 \\ \text{"i alt"} & 268 & 183 & 549 & 1000 \end{bmatrix}:$$

$bidrag(A)$

$$\begin{bmatrix} \text{"Bidrag"} & \text{"Aviser eller blade"} & \text{"Internet eller mobil"} & \text{"Radio eller TV"} & \text{"i alt"} \\ \text{"Kvinde"} & 0.01017255867 & 1.828685219 & 0.7245636744 & 2.563421452 \\ \text{"Mand"} & 0.01832200343 & 3.293682341 & 1.305026450 & 4.617030794 \\ \text{"i alt"} & 0.02849456210 & 5.122367560 & 2.029590124 & 7.180452246 \end{bmatrix} \quad (1.3)$$

Bidrag virker naturligvis også på matricere uden labels og summer

$$T := \begin{bmatrix} 171 & 103 & 369 \\ 97 & 80 & 180 \end{bmatrix}:$$

$bidrag(T)$

$$\begin{bmatrix} 0.01017255867 & 1.828685219 & 0.7245636744 \\ 0.01832200343 & 3.293682341 & 1.305026450 \end{bmatrix} \quad (1.4)$$

bidrag kan også benyttes i forbindelse med Goodness of Fit test. Her skal både den observede og forventede fordeling benyttes som input.

1.2 Regressioner

reg_koeff

er en global (beskyttet) variabel, der opdateres hver gang en regression (fra Gym-pakken) er udført. Fx

$X := [10, 13, 23, 35, 40, 50, 120]:$

$Y := [10.6, 8, 4.84, 4.36, 4, 2.25, 1.96]:$

$PowReg(X, Y, x)$

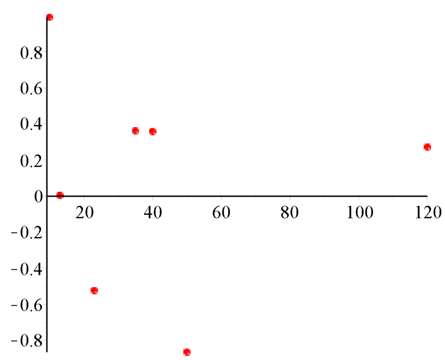
$$\frac{48.1305244985583}{x^{0.699783897538756}} \tag{1.5}$$

reg_koeff

$$\begin{bmatrix} b \\ a \end{bmatrix} = \begin{bmatrix} 48.1305244985583 \\ -0.699783897538756 \end{bmatrix} \tag{1.6}$$

residualer og plotResidualer

$plotResidualer(X, Y, PowReg)$



Ud over standardregressionerne i Gym-pakken, kan enhver modelfunktion f benyttes ved at bruge kommandoen $plotResidualer(X, Y, f)$.

Kommandoen *residualer* returnerer en en matrix med residualerne:

$residualer(X, Y, PowReg)$

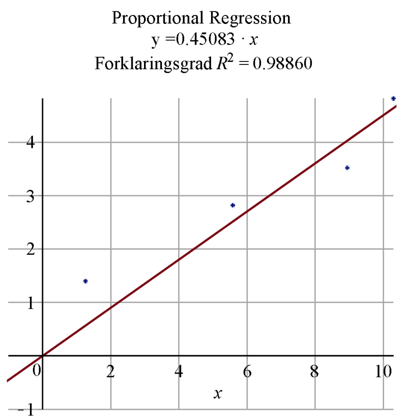
$$\begin{bmatrix} 10 & 0.991918089269102 \\ 13 & 0.00348099635898080 \\ 23 & -0.524205188951670 \\ 35 & 0.361412166268790 \\ 40 & 0.358126894032790 \\ 50 & -0.865363611981040 \\ 120 & 0.271725587520294 \end{bmatrix} \tag{1.7}$$

Proportional regression

$X := [1.26, 5.58, 8.94, 10.3]:$

$Y := [1.4, 2.82, 3.52, 4.82]:$

$PropReg(X, Y)$



reg_koeff

$$a = 0.4508342416 \tag{1.8}$$

1.3 Vektorregning

vsolve

- er en modifikation af Maple's solve-rutine, så vektorligninger tillades som input.

$\vec{a} := \langle 1, 2\rangle : \vec{b} := \langle -1, 3\rangle : \vec{c} := \langle -1, 13\rangle :$

$vsolve(s \cdot \vec{a} + t \cdot \vec{b} = \vec{c}, \{s, t\})$

$$\{s = 2, t = 3\} \tag{1.9}$$

Når elementerne er tal, kan parameterlisten udelades uden problemer:

$vsolve(s \cdot \vec{a} + t \cdot \vec{b} = \vec{c})$

$$\{s = 2, t = 3\} \tag{1.10}$$

Er vektorerne parametriserede, fx

$\vec{a} := \langle x, 2\rangle : \vec{b} := \langle -1, 3\rangle : \vec{c} := \langle -1, 13\rangle :$

giver vsolve uden parameterliste
$vsolve(s \cdot \vec{a} + t \cdot \vec{b} = \vec{c})$

$$\left\{s = \frac{10}{3x + 2}, t = \frac{13x + 2}{3x + 2}, x = x\right\} \tag{1.11}$$

og med parameterlisten

$vsolve\left(s\cdot\vec{a}+t\cdot\vec{b}=\vec{c},\ \{s,t\}\right)$

$$\left\{s=\frac{10}{3x+2},\ t=\frac{13x+2}{3x+2}\right\} \tag{1.12}$$

For 3D-vektorer fungerer det helt tilsvarende. Fx ka skæringslinjen mellem to planer (givet ved parameterfremstillinger) findes således:

$p := (s,t)\rightarrow\langle 0,-2,0\rangle+s\cdot\langle -4,0,4\rangle+t\cdot\langle 0,-2,4\rangle$:

$q := (u,v)\rightarrow\langle 0,0,3\rangle+u\cdot\langle -2,6,0\rangle+v\cdot\langle -2,0,3\rangle$:

$vsolve(p(s,t)=q(u,v))$

$$\left\{s=s,\ t=\frac{4}{3}+\frac{2}{3}s,\ u=-\frac{2}{9}s-\frac{7}{9},\ v=\frac{20}{9}s+\frac{7}{9}\right\} \tag{1.13}$$

Parameterfremstillingen for skæringslinjen findes ved at indsætte s og t i p - eller u og v i q:

$p\left(s,\frac{4}{3}+\frac{2}{3}s\right)$

$$\begin{bmatrix} -4s \\ -\frac{14}{3}-\frac{4}{3}s \\ \frac{20}{3}s+\frac{16}{3} \end{bmatrix} \tag{1.14}$$

1.4 Desriptiv statistik

antalobs

er en hjælpefunktion, der optæller antallet af observationer i et observationssæt - eller i en hyppighedstabel. Som input kan benyttes en liste, en vektor, en matrix eller et array. Hvis matricen er en hyppighedstabel, tilføjes hyppighed som parameter:

$X := [2,6,4,4,4,4,5,2,3,4,5,6,1]$:

$antalobs(X)=13$

$$H := \begin{bmatrix} 1 & 1 \\ 2 & 2 \\ 3 & 1 \\ 4 & 5 \\ 5 & 2 \\ 6 & 2 \end{bmatrix} :$$

$antalobs(H)=12$

$antalobs(H, hyppighed)=13$