Forløb om lineær regression afholdt i 2i-MA (mat A, samf A)

|  |  |
| --- | --- |
| Tidslinje (et modul er på 1,5 timer) | Aktivitet |
| 1. modul, mindste kvadraters metode | Lineær regression vises i Nspire på de fire datapunkter i eksemplet fra noterne ”regression\_og\_forklaringsgrad\_version5”. Dernæst vises residualkvadraterne med sum $SS\_{res}=1,8$. Tendenslinjen slettes og en flytbar linje med residualkvadrater introduceres.**Opgave**. Kan du få $SS\_{res}$ ned på 1,8?**Opgave**. Gør den flytbare linje vandret. Vis at minimummet for $SS\_{res}$ har $SS\_{tot}=26$. Vis at denne vandrette linje skærer *y*-aksen i gennemsnittet $\overbar{y}$. Beregn forklaringsgraden *r*2 vha. tallene 1,8 og 26.**Opgave**. Ekstraopgave 2 i noterne”regression\_og\_forklaringsgrad\_version5” hvor du beviser at $SS\_{res}$ er minimal når den vandrette linje har $b=\overbar{y}$.Med Nspire og NetLogo programmerne ”mindste\_kvadrater” vises det til sidst, hvordan *a* og *b* for tendenslinjen kan findes ved at steppe henimod minimummet for $SS\_{res}(a,b)$. |
| 2. modul$$a=\frac{sum(y∙\left(x-\overbar{x}\right))}{sum(\left(x-\overbar{x}\right)^{2})}$$$$b=\overbar{y}-a∙\overbar{x}$$$$r^{2}=1-\frac{SS\_{res}}{SS\_{tot}}$$$$SS\_{tot}=sum(\left(y-\overbar{y}\right)^{2})$$$$SS\_{res}=sum(\left(y-(a∙x+b)\right)^{2})$$$$\hat{σ}\_{res}=\sqrt{\frac{sum(res^{2})}{n-2}}$$$$res\_{i}=y\_{i}-(a∙x\_{i}+b)$$ | Lineær regression vises på data i solfangereksemplet. Her er *x* opvarmningstiden i minutter og *y* er temperaturen i °C.**Opgave**. Vis at formlerne for *a*, *b* og *r*2 giver Nspires værdier for lineær regression på data i solfangereksemplet.Det vises hvordan residualspredningen $\hat{σ}\_{res}$ bestemmes i beregneren og i regnearket som outputtet *s* fra ”statistik -> konfidensintervaller -> lineær reg t-intervaller”. Dernæst vises en normalfordelingstest af residualerne i et ”diagrammer og statistik” vindue. Den bagvedliggende teori om standardnormalfordelingen af $z=\frac{x-μ}{σ}$ repeteres.Det opklares at residualspredningen $\hat{σ}\_{res}$ er relateret til spredningen i normalfordelingstesten ved $\hat{σ}\_{res}=\sqrt{\frac{n-1}{n-2}}∙\hat{σ}\_{normtest}$.**Opgaver** fra bogen. |
| 3. modul$$a\in [CLower, CUpper]$$$$a=a\_{0}\pm ME$$ | Elevgennemgang af bogens opgaver fra sidste gang.NetLogo programmet ”fordeling\_a\_b\_y” introduceres. Alle *y*’er er normalfordelte omkring den sande linje. Det vises at der opnås en pæn fordeling for en variabel *ta* der er knyttet til hældningen *a*. Den præcise definition af *ta* gemmes til næste modul. 95% konfidensintervallet for hældningen *a* bestemmes for data i solfangereksemplet vha. ”statistik -> konfidensintervaller -> lineær reg t-intervaller”.**Opgaver** fra bogen. |
| 4. modul$$t\_{a}=\frac{a-a\_{sand}}{\hat{σ}\_{a}}$$$$\hat{σ}\_{a}=\hat{σ}\_{res}∙\sqrt{\frac{1}{sum(\left(x-\overbar{x}\right)^{2})}}$$$$t\_{b}=\frac{b-b\_{sand}}{\hat{σ}\_{b}}$$$$\hat{σ}\_{b}=\hat{σ}\_{res}∙\sqrt{\frac{1}{n}+\frac{\overbar{x}^{2}}{sum(\left(x-\overbar{x}\right)^{2})}}$$ | Elevgennemgang af bogens opgaver fra sidste gang.**Opgave**. Lav arbejdsarket ”t-fordelinger\_opgaver”.NetLogo programmet ”fordeling\_a\_b\_y” introduceres, men denne gang med formler for *ta* og *tb*.**Opgave**. Undersøg vha. programmet ”fordeling\_a\_b\_y” om fordelingerne for *ta* og *tb* er t-fordelinger med *n* - 2 frihedsgrader.Det vises hvordan man selv kan beregne tallene i outputtet fra ”statistik -> konfidensintervaller -> lineær reg t-intervaller”.Det gennemgås at en ikke-sammenhæng følger fra $a=0$ eller $a\in [negativt tal, positivt tal]$.**Opgaver** fra bogen. |
| 5. modul, test for $a=0$ | Elevgennemgang af bogens opgaver fra sidste gang.Der vises en test for$ a=0$ på data i ”Regn i Voer” eksemplet. Her er *x* tiden i år efter 2009 og *y* er regnmængden i mm per år. Der anvendes en tosidet ”statistik -> statistiske tests -> lineær regressions t-test”.Dernæst gives der en forklaring på hvad der ligger bagved testen og forbindelsen til 95% konfidensintervallet for hældningen *a*.**Opgaver** fra bogen. |