

Eksperiment vedrørende residualspreddningen

I kapitel 2 fra bog 1 arbejdede vi med spredningen af n data i en population eller en stikprøve. Vi kan med et simulationseksperiment argumentere for, at spredningen for en stikprøve giver et bedre estimat på populationens spredning, hvis vi i udtrykket for stikprøvespredningen dividerer med $n-1$ og ikke n . Vi kan med samme strategi argumentere for, at udtrykket for residualspreddningen giver et bedre estimat for populationens residualspreddning, hvis vi dividerer med $n-2$ og ikke n .

Eksperimentet er opbygget efter følgende strategi:

- Vælg 10 forskellige x -værdier i et regneark.
- Vælg en lineær funktion $f(x) = ax + b$. Dvs. vælg en værdi for a , og vælg en værdi for b .
- Bestem på baggrund af den lineære funktion f , de tilhørende 10 funktionsværdier i regnearket.
- Bestem 10 støjled, der er normalfordelt med middelværdi 0 og spredning 1. De 10 støjled skal udregnes med en tilfældighedsgenerator i regnearket.
- Læg det tilhørende støjled til hvert af de 10 funktionsværdier, så vi får 10 nye tilfældige funktionsværdier i regnearket.
- Bestem på baggrund af de sammenhørende x -værdier og tilfældige funktionsværdier en lineær regressionsmodel g .
- Bestem de 10 residualer i regnearket. Vi kalder dem r_1, r_2, \dots, r_{10} .
- Bestem $\hat{\sigma}_{bud1} = \sqrt{\frac{r_1^2 + r_2^2 + \dots + r_{10}^2}{10}}$ i regnearket.
- Bestem $\hat{\sigma}_{bud2} = \sqrt{\frac{r_1^2 + r_2^2 + \dots + r_{10}^2}{9}}$ i regnearket.
- Bestem $\hat{\sigma}_{bud3} = \sqrt{\frac{r_1^2 + r_2^2 + \dots + r_{10}^2}{8}}$ i regnearket.
- Bestem $\hat{\sigma}_{bud4} = \sqrt{\frac{r_1^2 + r_2^2 + \dots + r_{10}^2}{7}}$ i regnearket.
- Bestem $\hat{\sigma}_{bud5} = \sqrt{\frac{r_1^2 + r_2^2 + \dots + r_{10}^2}{6}}$ i regnearket.
- Gennemfør dette eksperiment 1000 gange, og opsaml værdierne for $\hat{\sigma}_{bud1}$, $\hat{\sigma}_{bud2}$, $\hat{\sigma}_{bud3}$, $\hat{\sigma}_{bud4}$ og $\hat{\sigma}_{bud5}$ i regnearket.
- Bestem ud fra de 1000 værdierne for $\hat{\sigma}_{bud1}$, $\hat{\sigma}_{bud2}$, $\hat{\sigma}_{bud3}$, $\hat{\sigma}_{bud4}$ og $\hat{\sigma}_{bud5}$ 5 forskellige gennemsnit, og vurder hvilket af de 5 gennemsnit, der rammer $\sigma = 1$ bedst.

Bemærkning: I eksperimentet kunne vi vælge andre værdier for a , b og σ . Fx kunne vi blot vælge værdierne $a = 0$, $b = 0$ og $\sigma = 1$.