Arbejdsseddel: Regressionsanalyse

KBJ, november 2023 1s Ma

**Opgave 1**

1. Hent på Lectio filen ”Data.xlsx”. Konvertér decimal-komma til decimal-punktum i begge datarækker og kopiér data til Nspire. Kald datalisterne for ”xdata” og ”ydata”.
2. Lav ny side med ”diagrammer og statistik”-vindue. Sæt ”xdata” på førsteaksen og ”ydata” på andenaksen. Udfør *lineær regression* på data (vælg ”$mx+b$”-varianten).
Sørg for at bedste rette linjes ligning vises, samt at man kan se værdien af $r^{2}$ (afkryds ”Diagnosticering” under ”Indstillinger” i Værktøjs-menuen).
3. Lav ny side med ”diagrammer og statistik”-vindue. Sæt ”datax” på førsteaksen og ”stat.resid” på andenaksen. Diskutér hvad der kan konkluderes ud fra residualplottet.
4. Brug residualplottet til at bestemme residualet for observationen med $x=24$.

Så snart en regression er udført, oprettes en liste med residualerne, som kaldes ”stat.resid”. Listen hentes frem i Nspire med at skrive ”stat.” og derpå vælge ”resid” på den fremkomne menu.

1. Lav et ”beregninger”-vindue, og bestem største og mindste værdi i stat.resid-listen. Brug kommandoer fra noten *Deskriptiv statistik med Nspire*. Hvad er den største afvigelse mellem en observation og ”bedste rette linje”?
2. Definér den tilhørende lineære model som: f(x):=*stat.m*·x+*stat.b*, hvor ”stat.m” og ”stat.b” hentes på samme måde, som ”stat.resid”.
3. Bestem den *forventede* modelværdi for $x=20$ ved at bestemme $f\left(20\right)$.
4. Bestem for hvilken $x$-værdi der *forventes* $y=370$ ved at løse ligningen $f\left(x\right)=370$.
5. Aflæs via punktplottet eller tabellen med data, den *observerede* værdi hørende til $x=47$.
Beregn med modellen den tilhørende *forventede* modelværdi.
Bestem den *absolutte afvigelse* (residualet) mellem *observeret* og *forventet* værdi, for $x=47$.
Bestem den tilhørende *relative afvigelse* (i procent) mellem *observeret* og *forventet* værdi.
6. Bestem for hvilken $x$-værdi der er fundet det største residual.
7. Beregn den relative afvigelse mellem *observeret* og *forventet* værdi, for det punkt i datasættet, der har det største residual.