

Skizofreni er en alvorlig og sjælden sindssygdom, som præger evnen til at tænke og føle, samt at skelne virkelighed og fantasi fra hinanden. Der findes fem forskellige typer af skizofreni, og heraf er paranoid skizofreni den mest almindelige. Symptomer omfatter vrangforestillinger, tankeforstyrrelser, nedsat evne til koncentration samt hallucinationer mm.

Behandlingen af skizofreni er et længere forløb, som kan indebære medicin, støttende samtaler og undervisning i sygdommen samt meget mere. Omkring 20 procent af de ramte bliver raske igen.

Trods det er der stadig meget man ikke ved om sygdommen, selvom forskningen fortsætter.

Mænd og kvinder

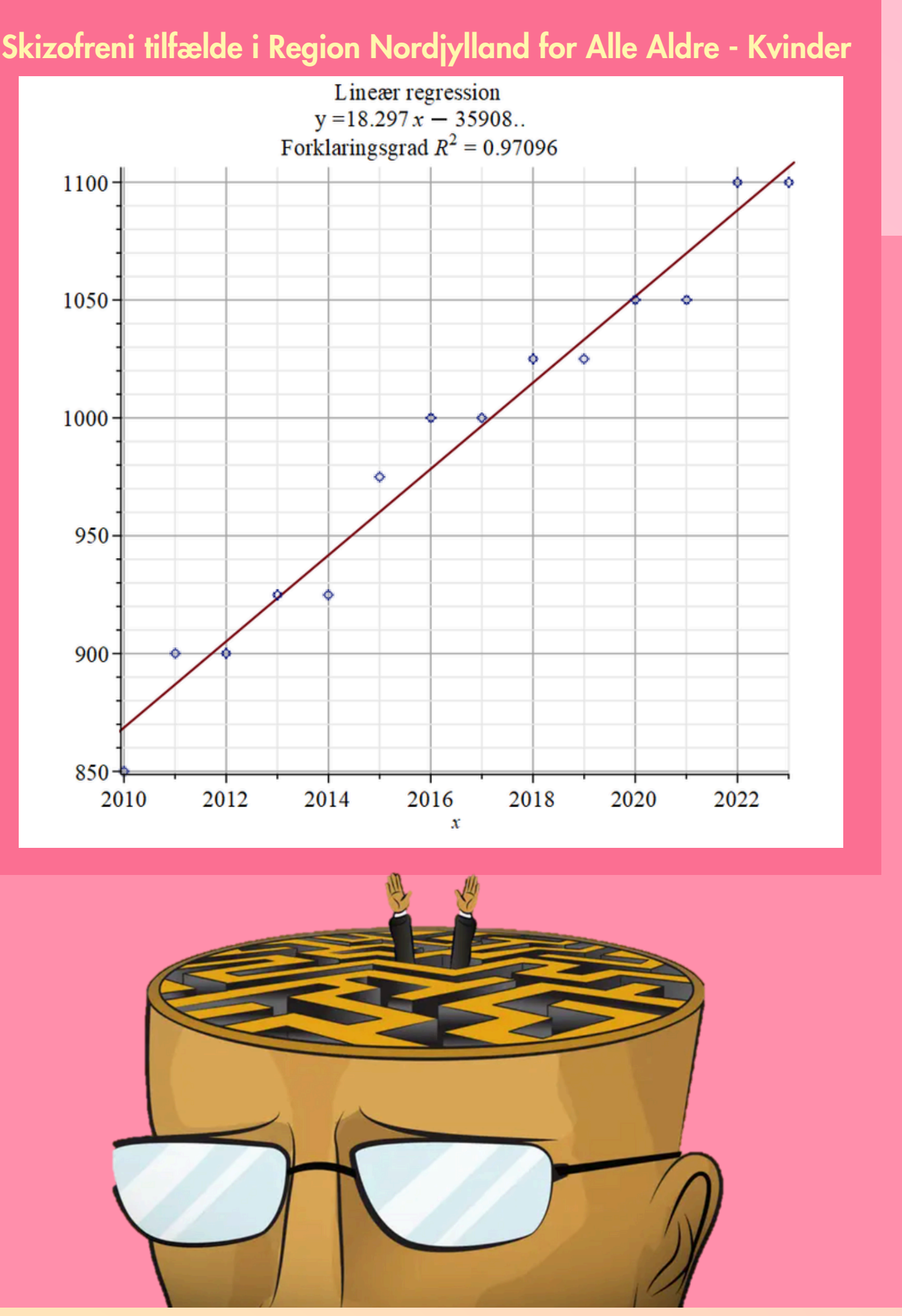
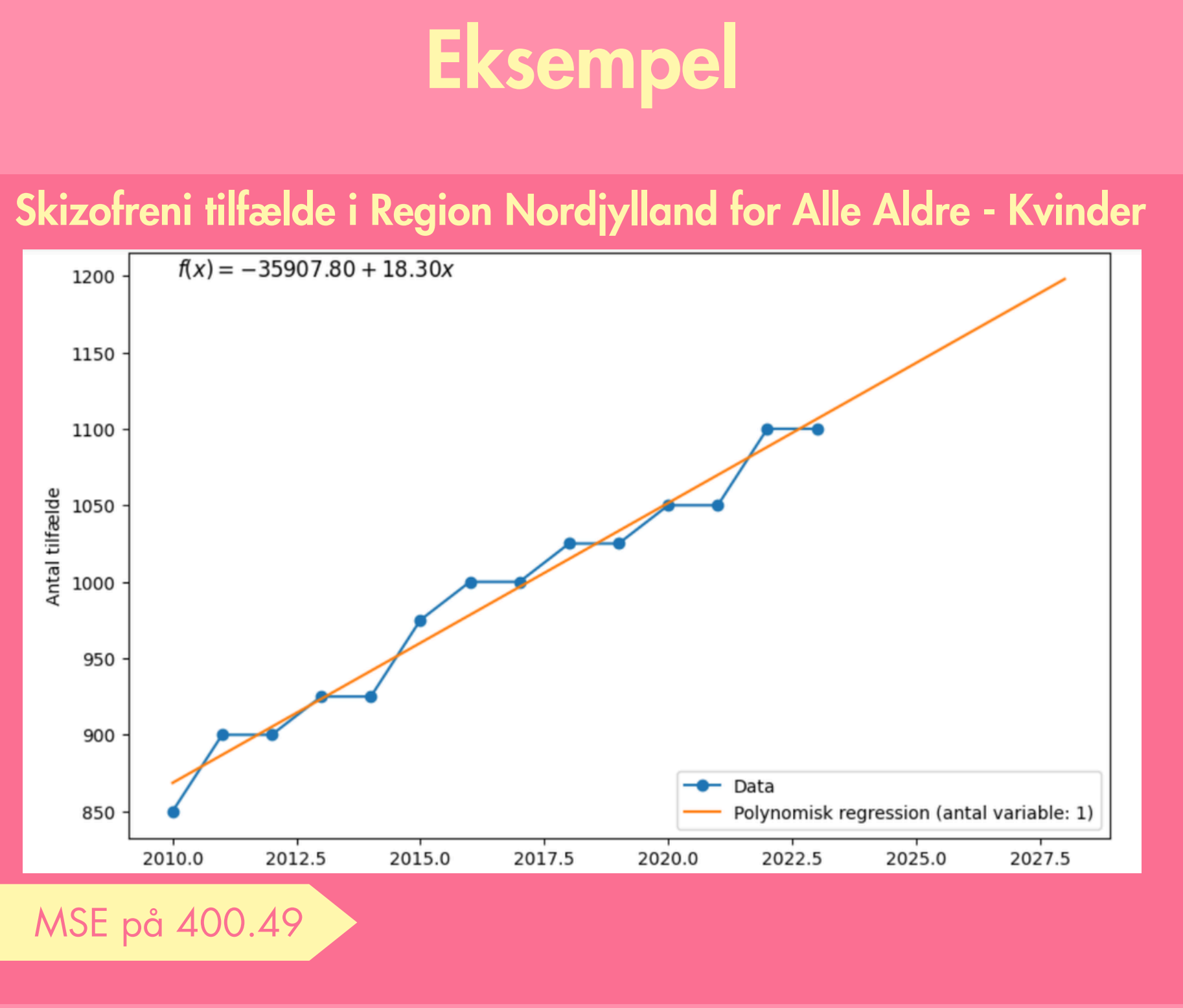
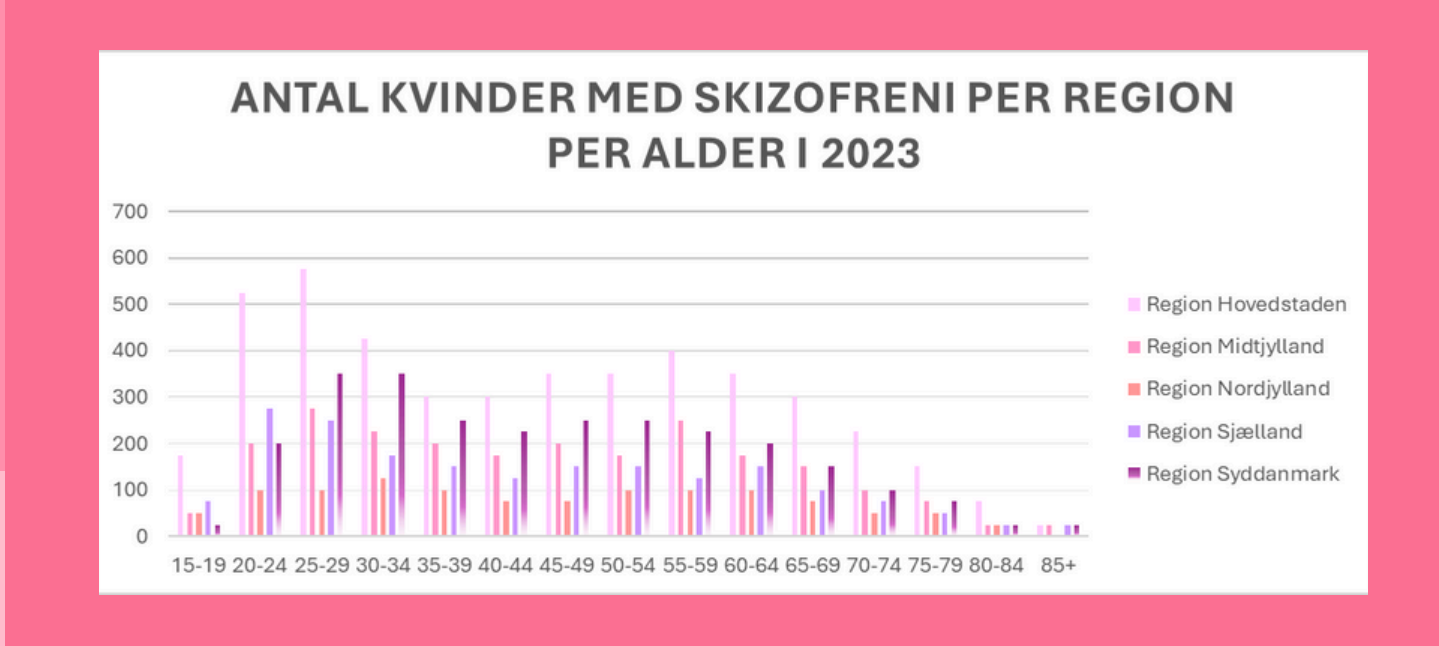
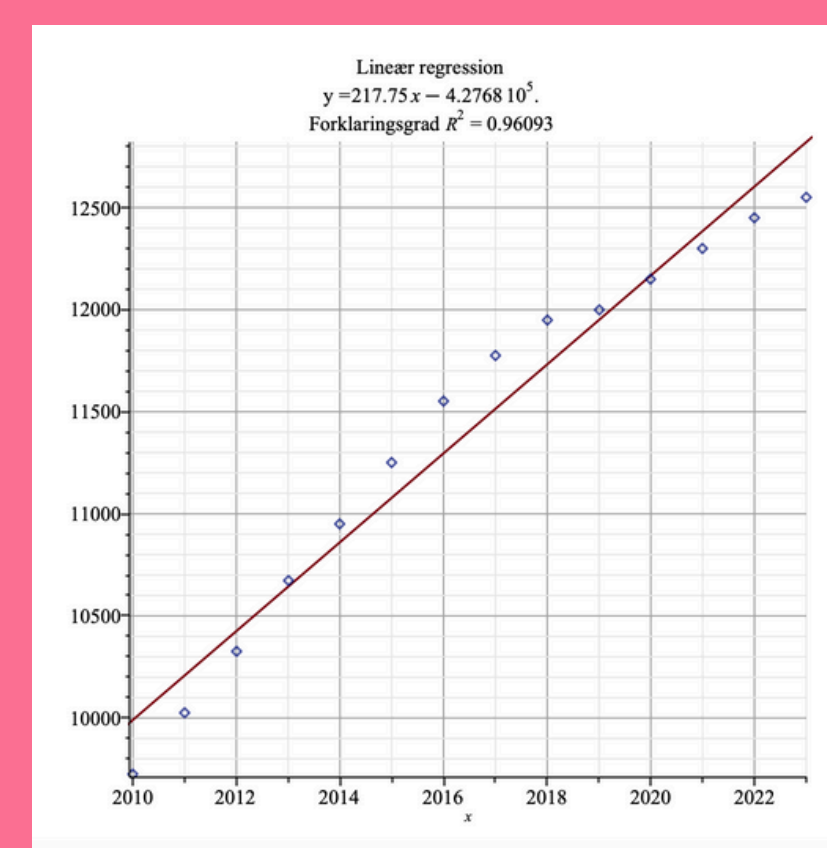
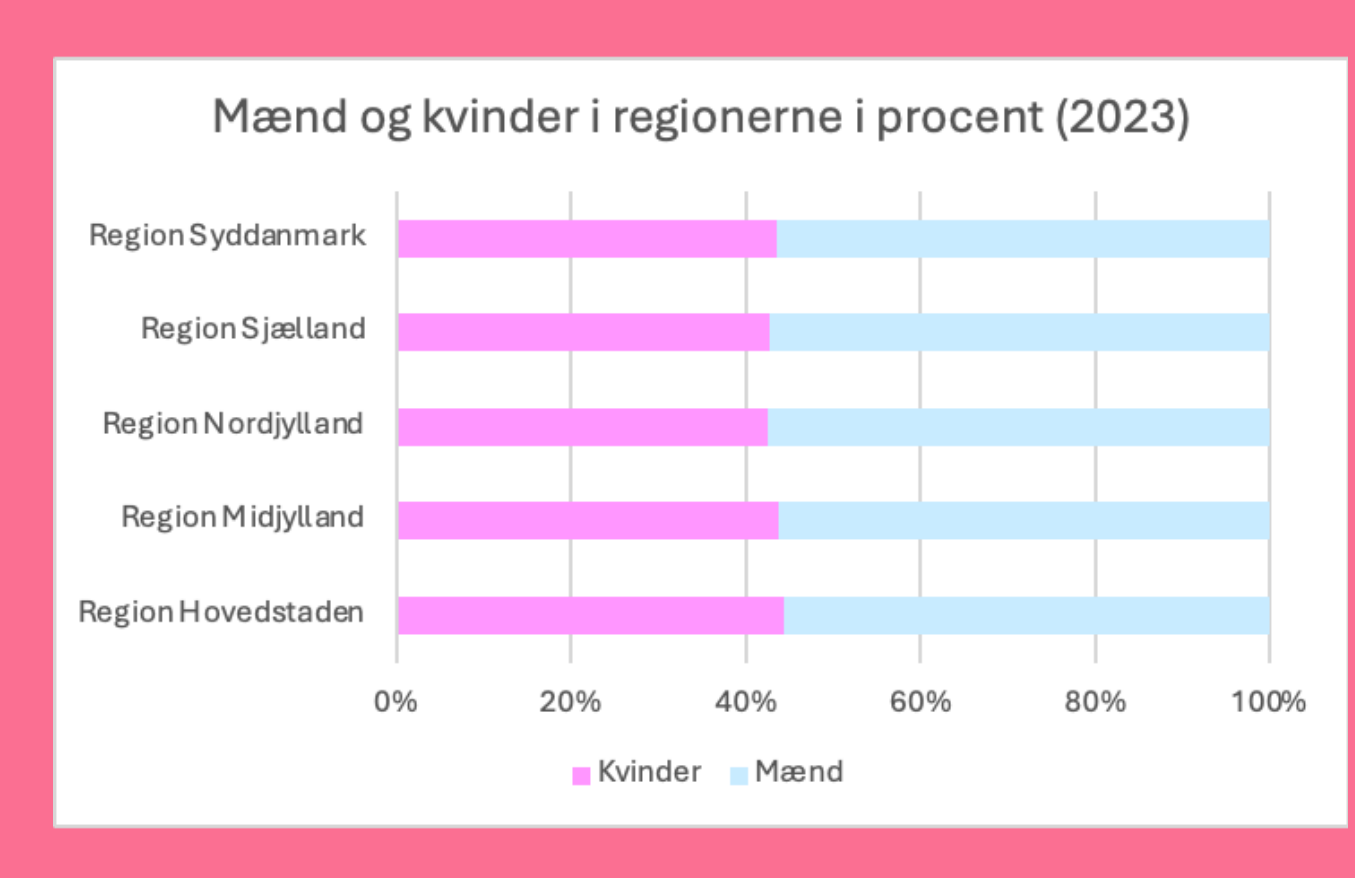
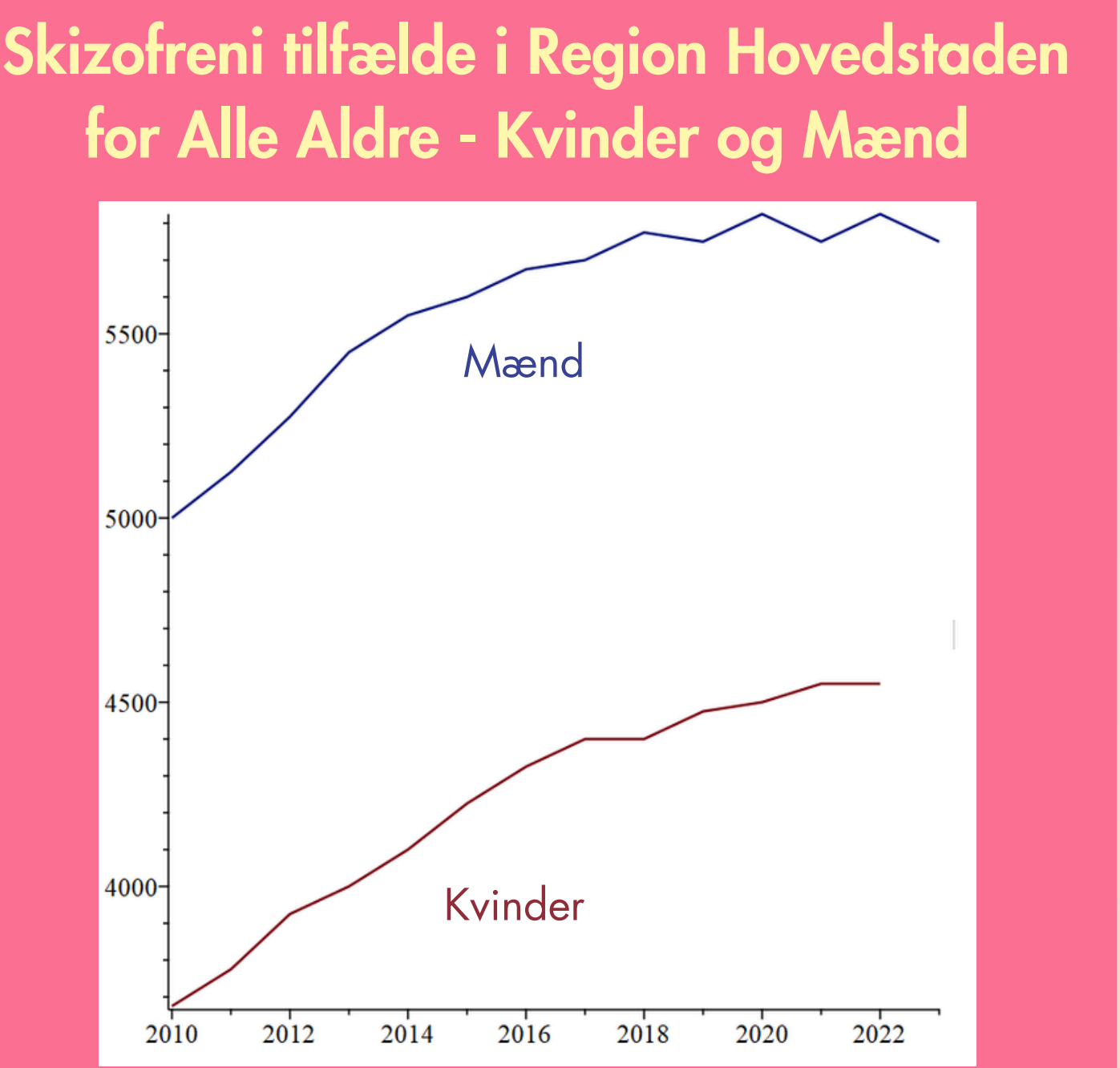
Vi har lagt vores fokus på den kvindelige befolkning, da mænd har større chance for at blive skizofrene end kvinder. Livsrisikoen for mænd er cirka 2%, mens den for kvinder “kun” er 1,5%. Dog er der 40% flere mænd der får stillet diagnosen end kvinder ifølge psykologeridanmark.dk, men det interessante er at vores data viser at der næsten lige mange mænd og kvinder i danmark som får stillet i diagnosen. Dette betyder, at flere kvinder end mænd går udiagnosticeret, hvilket er et kæmpe problem.

Data

Vores datasæt indeholder information om alle i danmark med skizofreni fordelt på region, alder og køn. Vi valgte at fokusere på kvinder og på forskellen mellem mænd og kvinder.

Histogrammet under viser antallet af skizofrene kvinder i 2023. Der er flest kvinder med skizofreni i alderen 25-29 år og i alle aldre er der flest skizofrene kvinder i Hovedstaden. Grafen til højre viser stigningen af kvinder i hele Danmark med skizofreni.

Denne graf viser stigningen i både mænd og kvinder med skizofreni i alle regioner. Hovedstaden er den region med flest skizofrene og størst stigning, men er samtidig også regionen med flest mennesker. Region midtjylland har derimod den mindste stigning.



Skizofreni

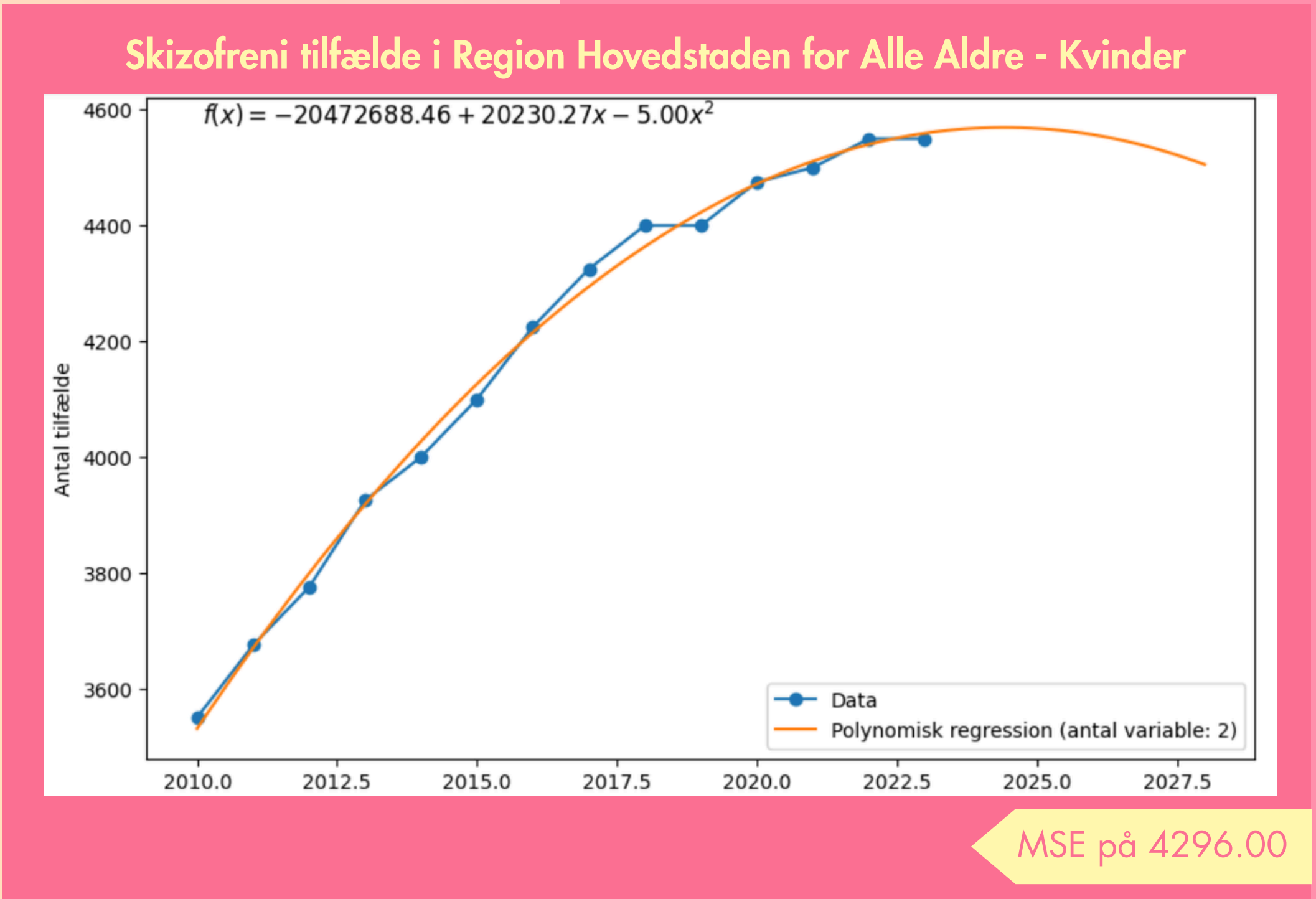
Overfitting og krydsvalidering

Hvordan kan man vælge den bedste model til sine data? Et mål for hvor tæt ens polynomium er på ens data er kvadratsummen af afvigelserne på modellen, hvilket beregnes ved:

$$E = \sum_{i=1}^n (y_i - f(x_i))^2$$

Et tredjegradspolynomium vil altid passe bedre til dataet end et andengradspolynomium, hvilket skyldes, hvis man sætter $a_3 = 0$, kommer der et andengradspolynomium. Derfor kunne man tænke, at jo højere en grad, jo bedre en model. Kvadratsummen for en model med en højere grad, vil ofte være bedre, da grafen prøver at følge datapunkterne meget tæt. Det største problem opstår derfor, hvis man kommer med nye datapunkter - dette fænomen kaldes altså overfitting. I en sådan model passer den virkelig godt til de kendte datapunkter, men ville være særdeles dårlig til at forudsige nye datapunkter.

Krydsvalidering er en måde hvorpå, man kan mindske risikoen for overfitting. Her gentager man trænings- og testproceduren flere gange. Det kaldes k -fold krydsvalidering, og der deles i k lige store og tilfældige dele. Herefter trænes modellen på alt data bortset fra første del - tabsfunktionen $E_1(p)$. Denne procedure gentages k gange. Summen af tabsfunktionerne fra de k fold bruges som et mål for, hvor god modellen er. Modellen der giver den mindste værdi af $E(p)$ er bedst.



$$E = \sum_{i=1}^n (y_i - (a_0 + a_1x_i))^2$$

Formel for Mean Squared Error. Her er listen udtrykt ved (x_i, y_i) , hvor 'i' er en talrække fra 1-n. Derudover er 'a₀' linjens skæring med y-aksen, og 'a₁' er hældningen.

Konklusioner og sammenhænge

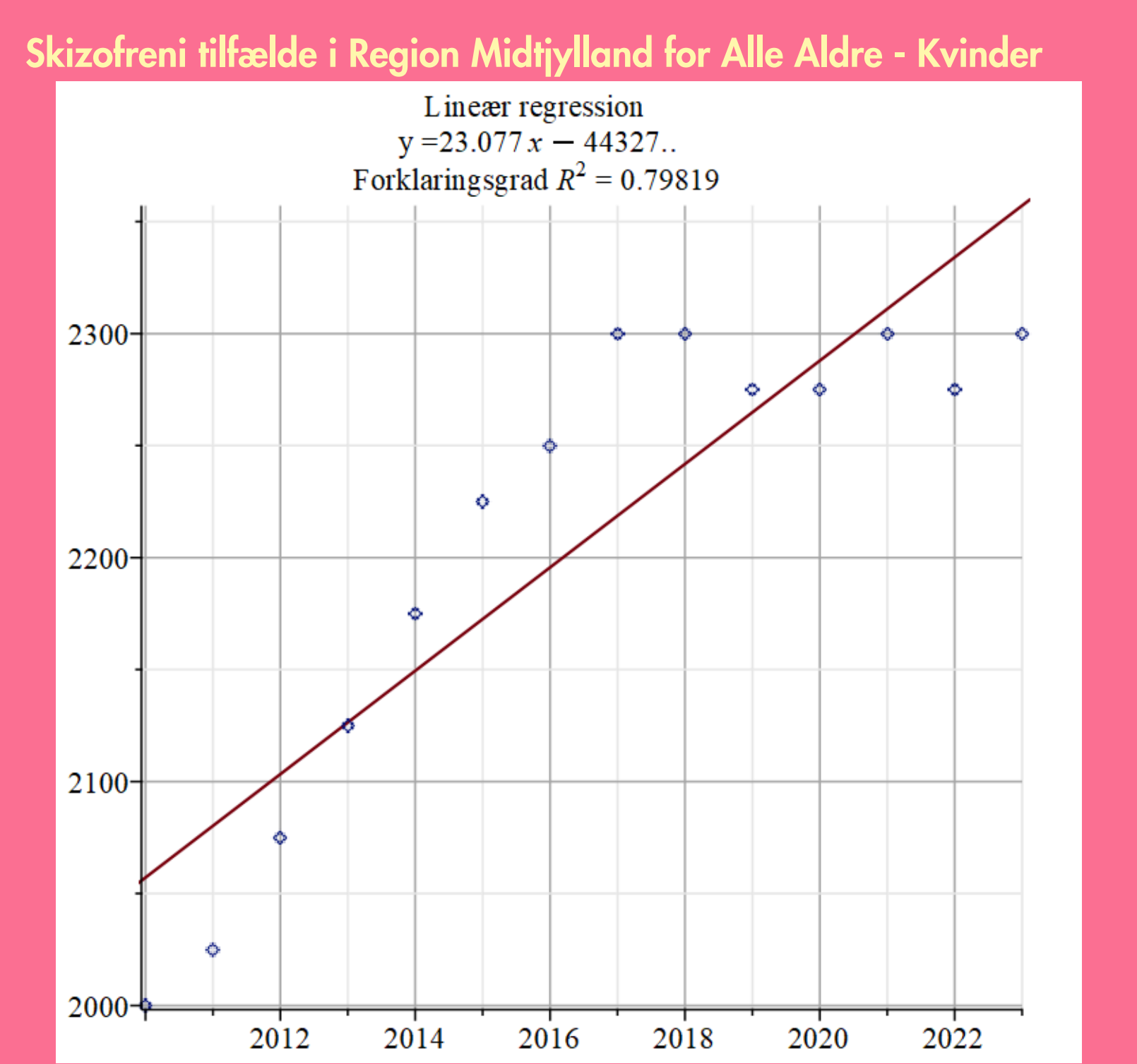
Vi kan konkludere på baggrund af vores deskriptive data, at der generelt er flere mænd end kvinder, der er skizofrene. Generelt for kvinderne er der flest med skizofreni i Region Hovedstaden, hvilket giver mening, da der bor flest i Hovedstaden ift. de andre regioner. Derudover er det hyppigst for kvinder i alderen 25-29 år. Der kan også konkluderes, at antallet af skizofrene vokser for hvert år, hvilket tyder på, at vi bliver bedre til at diagnosticere.

Ved brug af metoden krydsvalidering, hvor der trænes på 4/5 af dataet, bliver de polynomiske regressioner af flere grader brugt til at beskrive dataet bedst muligt, dog konkluderes der, at modellerne kun kan forudsige indenfor kort fremtid.

Polynomisk regression

For at finde ud af hvilken model, der bedst kan beskrive udviklingen af skizofreni i befolkningen, kan man bruge regression. Det simpleste eksempel på regression er lineær. Ved lineær regression finder man den bedste rette linje til datasættet.

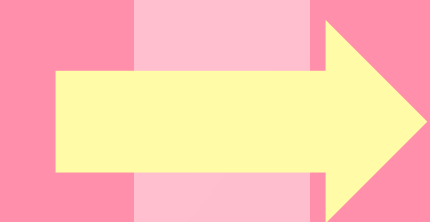
Her ses en lineær regression af skizofreni tilfælde i region Midtjylland for alle aldre i alle år hos kvinderne. Her kan det ligges mærke til at punkterne ikke passer helt godt til grafen. Nogle gange er det fordi, at dataene ikke følger en lineær tendens. Det kunne f.eks. være polynomisk regression. En polynomisk regression er en regression med en grad større end 1. Polynomisk regression er god, når dataene 'skifter retning' altså går fra at stige i antal per periode til at falde i antal per periode.



Mean Squared Error

Her ses en polynomisk regression af skizofreni tilfælde i region Hovedstaden for alle aldre i alle år hos kvinderne. Polynomisk giver en bedre tilpasning til data, men dette kan også føre til overfitting. Derfor bruger man metoden "Mean Squared Error" (MSE). Det bruges til at beregne nøjagtigheden af ens model, da polynomisk regression tilpasser sig rigtig godt til kendt data, men kan have en dårligere prædiktationsevne. Specifikt beregner MSE den gennemsnitlige kvadrerede forskel mellem de forudsagte værdier fra modellen og de rigtige værdier. På eksemplet ses en polynomisk regression, som har en MSE på 4296.00. MSE kan ved første øjekast virke meget høj, men her er det vigtigt at huske på, at MSE er den kvadrede forskel. Det betyder, at forskellene ofte bliver meget store, selvom der er god sammenhæng mellem graf og data.

Forventninger og anbefalinger



Der er klart flest skizofrene i hovedstaden, og flere mænd end kvinder har skizofreni. Dog er tallet for både mænd og kvinder stigene i hovedstaden.

I følge vores polynomiske regression vil antallet af skizofrene i hovedstaden falde, dog er der en del usikkerhed i modellen, da den er bedst til at forudsige punkter tæt på det kendte datasæt.

