

Á eftirfarandi glærum eru myndir úr:

- Bókinni "An Introduction to Statistical Learning" eftir James, Witten, Hastie og Tibshirani, 2. útg. 2021, ("SL-bók", sjá statlearning.com)
- Grein um ofnæmi barna í tímaritinu Allergy (Risk factors of school age allergy, sjá [PDF](#))

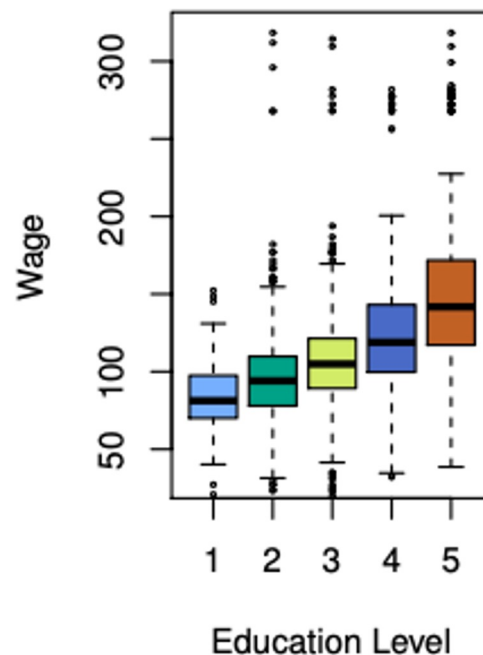
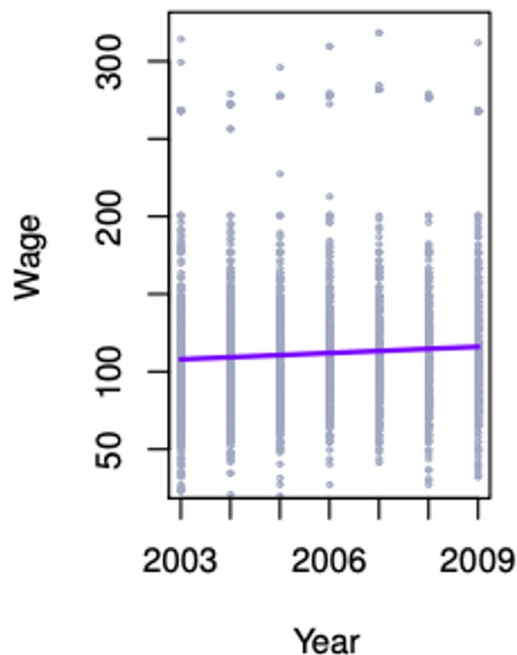
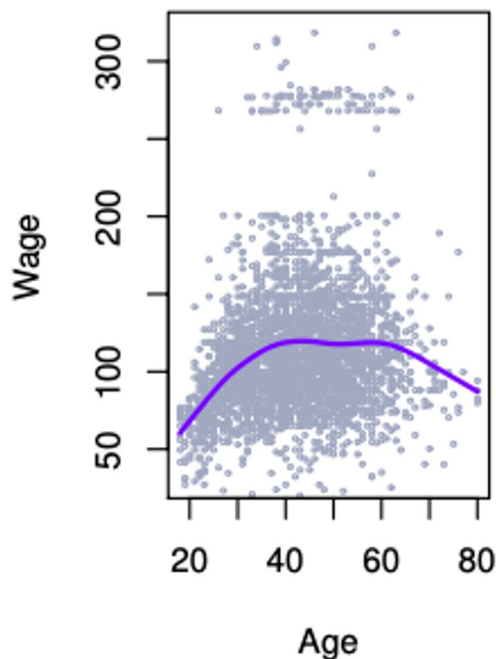
Í SL-bókinni er umfjöllun um tvíkostaaðhvarfsgreiningu (logistic regression).

Auk þess skoðum við umfjöllun um "logistic regression" á:

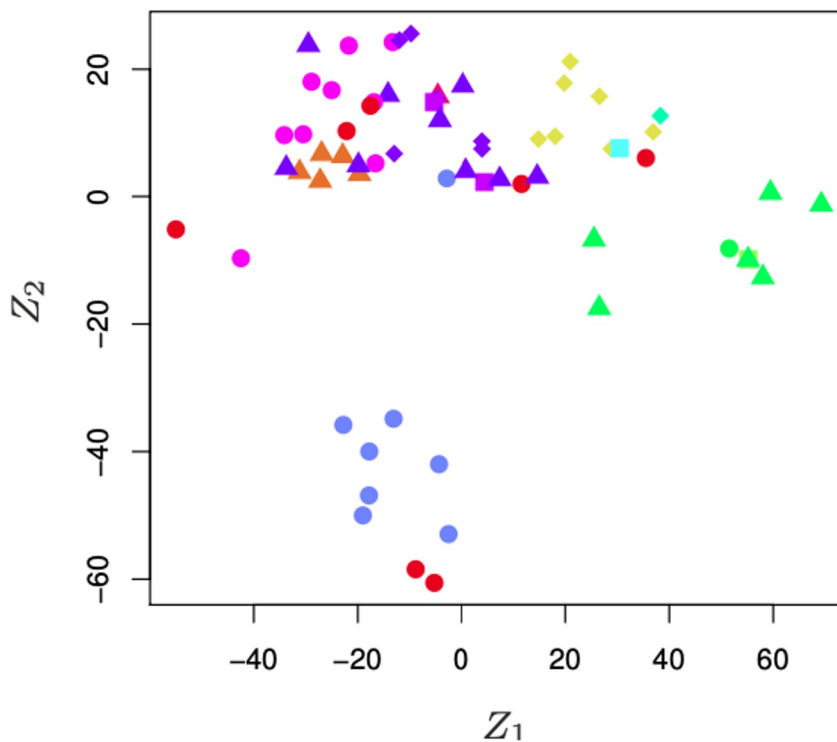
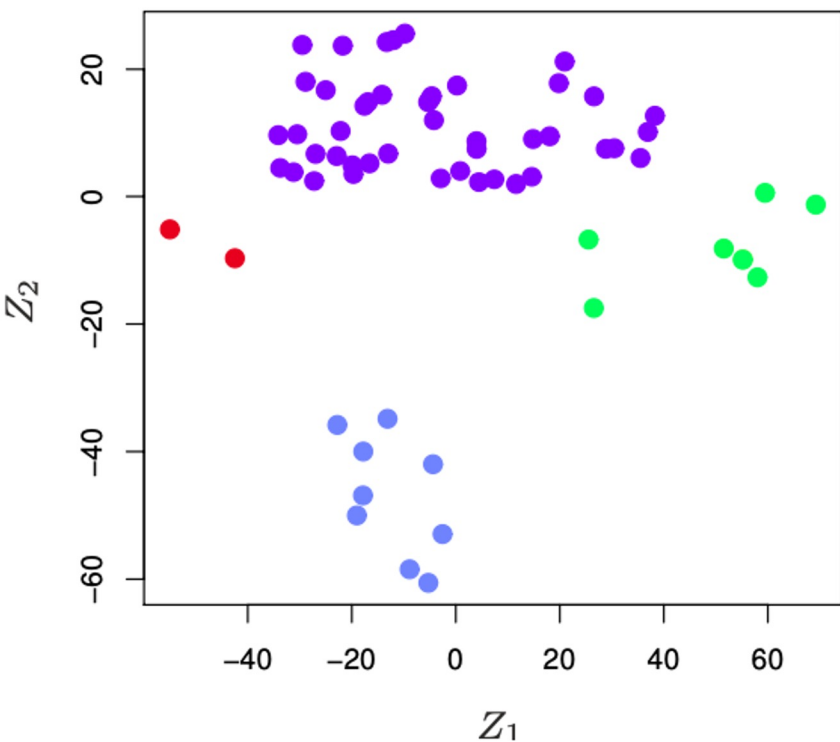
- [Wikipedíu](#)
- [Edbook-vefnum](#) (Tölfræði frá grunni)

Launagögn úr SL-bók

Gögn um laun karla á austurströnd bandaríkjanna (\$1000/ár). ESL = "Elements of statistical learning"



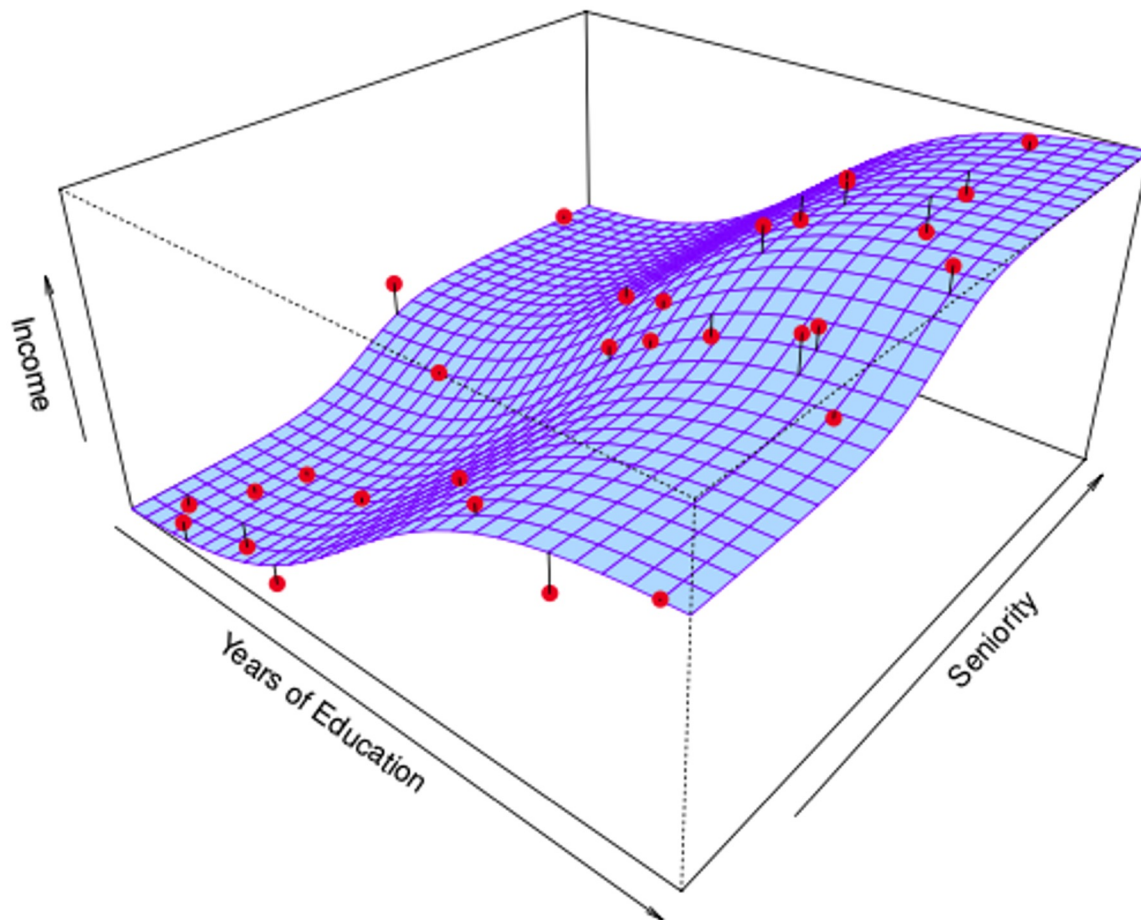
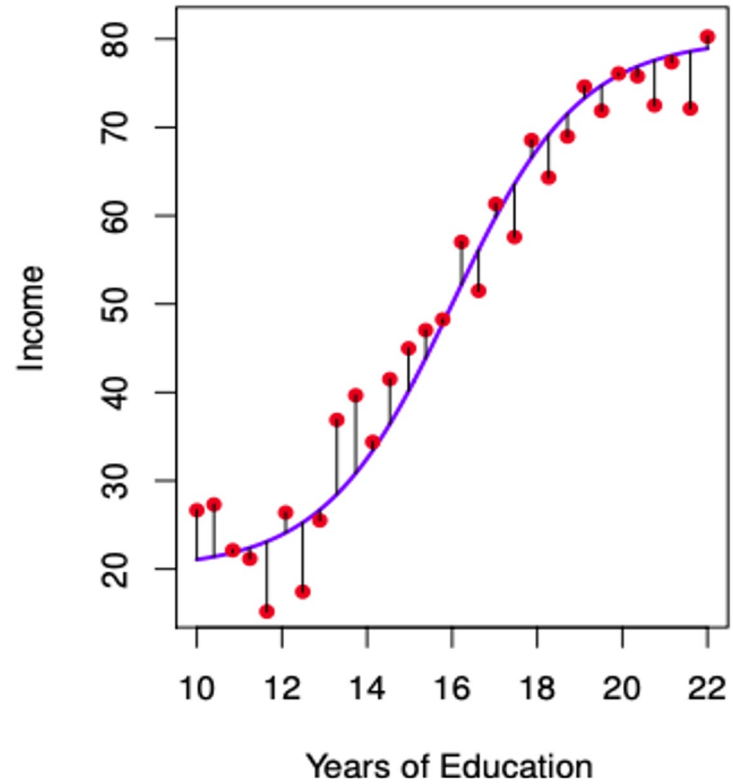
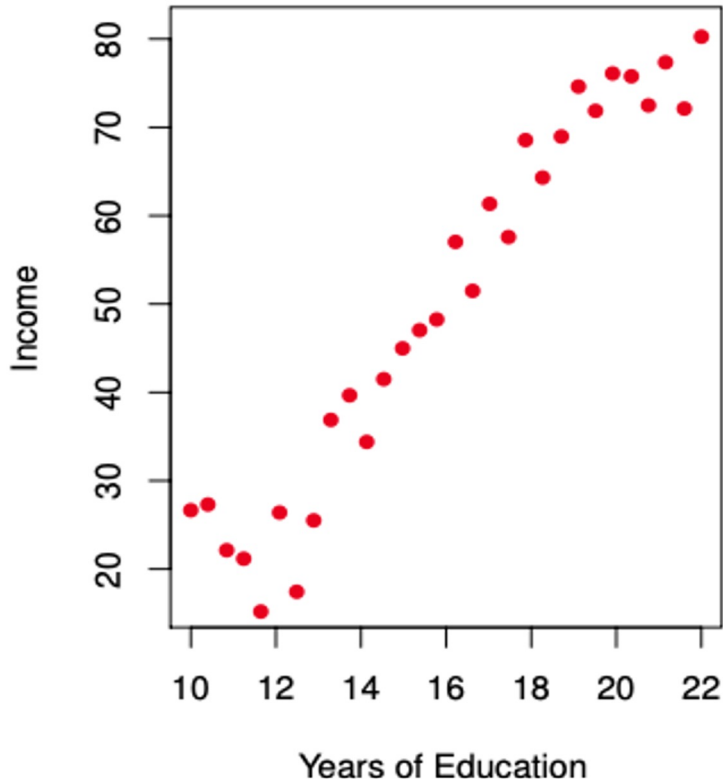
Genatjáníng



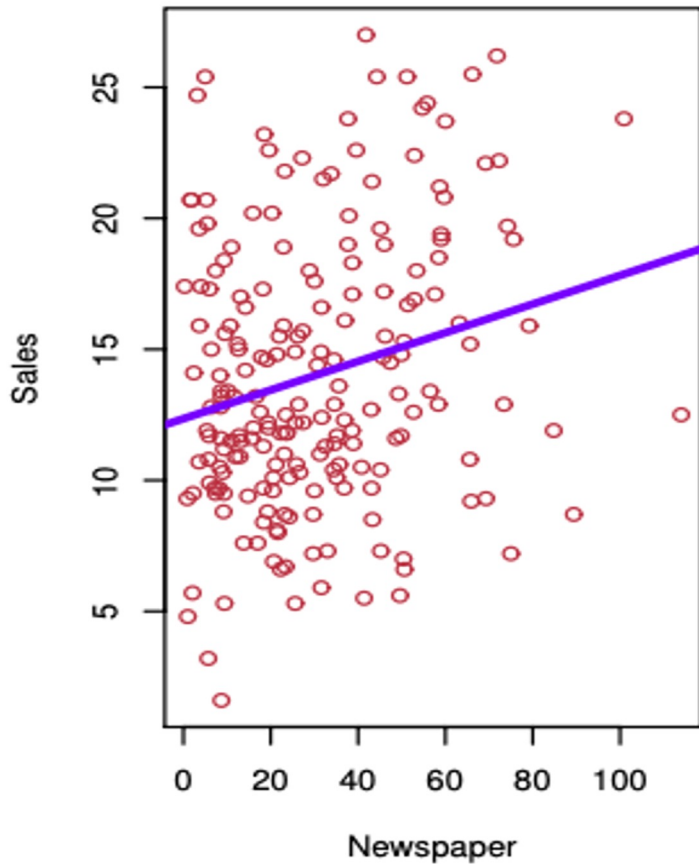
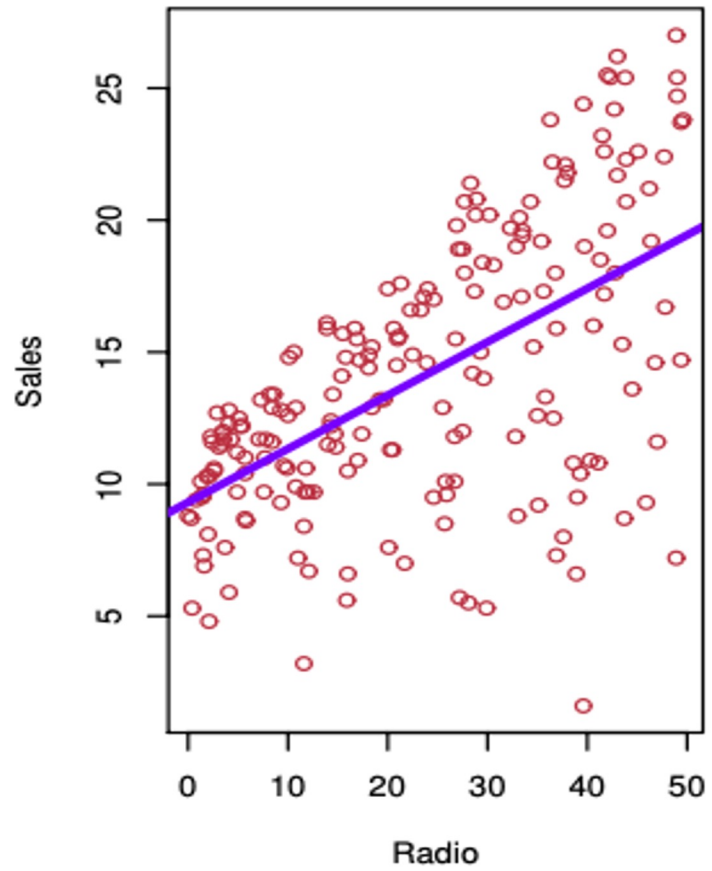
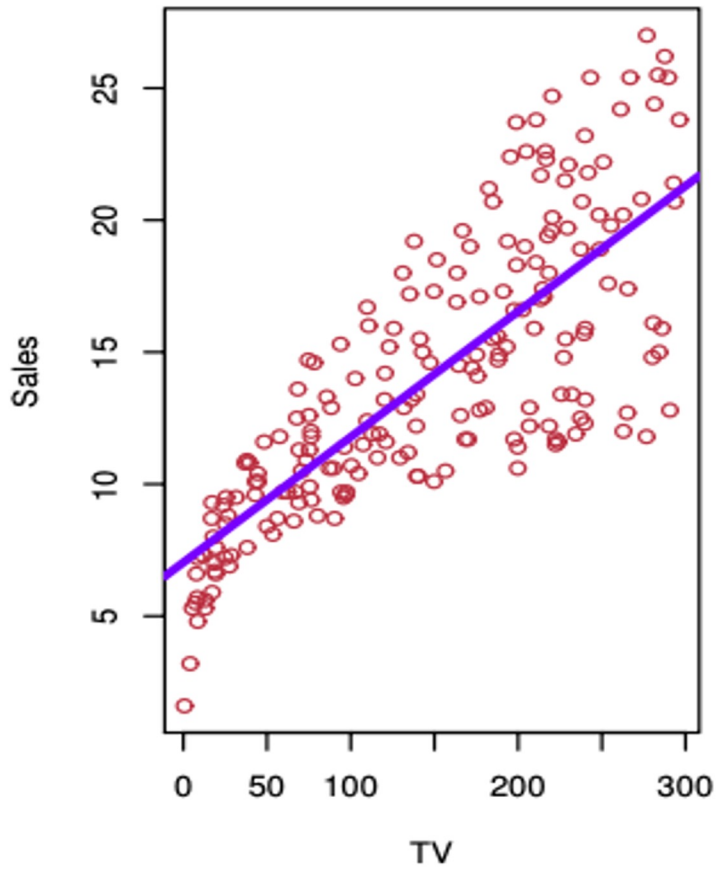
Gögn um genatjáníngu 14 krabbameina. Vinstra megin er k-means flokkun

Hermd launagögn

Bláa kúrfan og flöturinn sýna sambandið sem var notað til að herma.

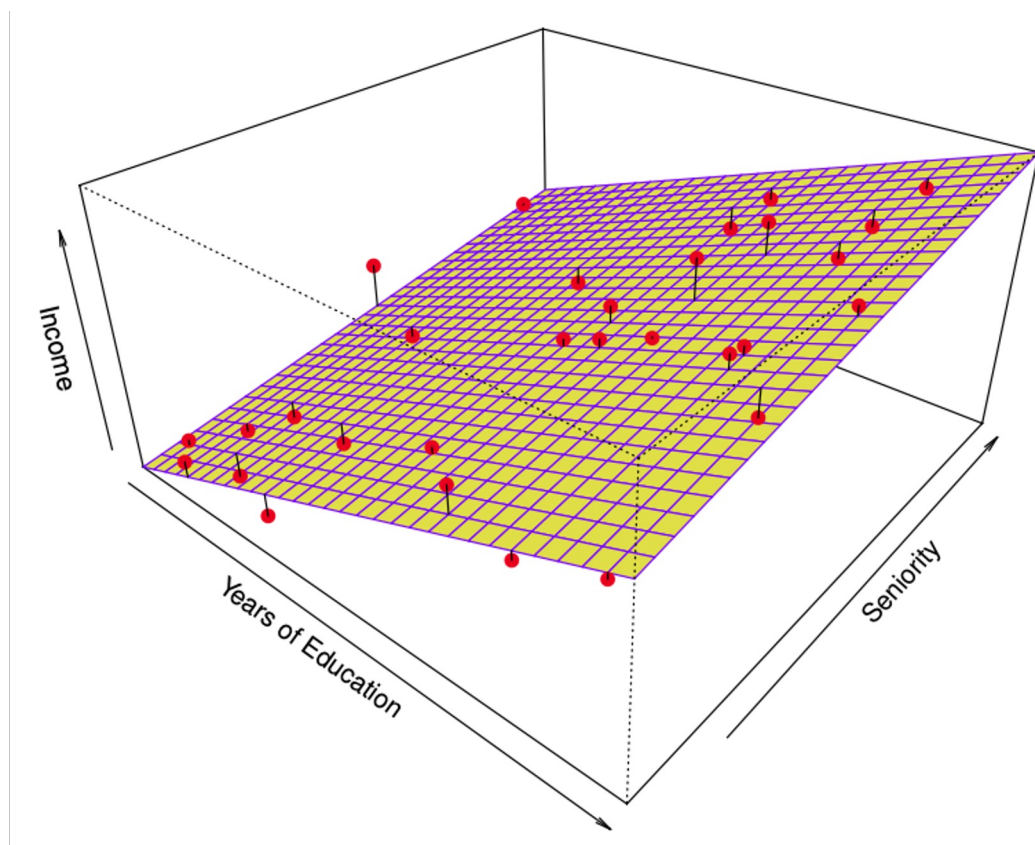


Auglýsingagögn

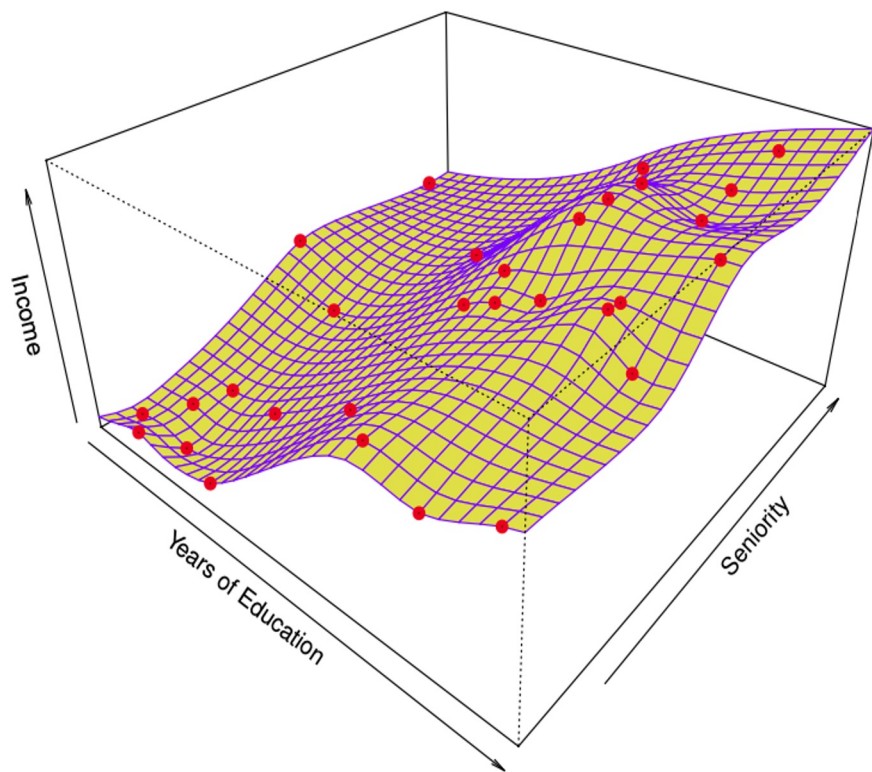
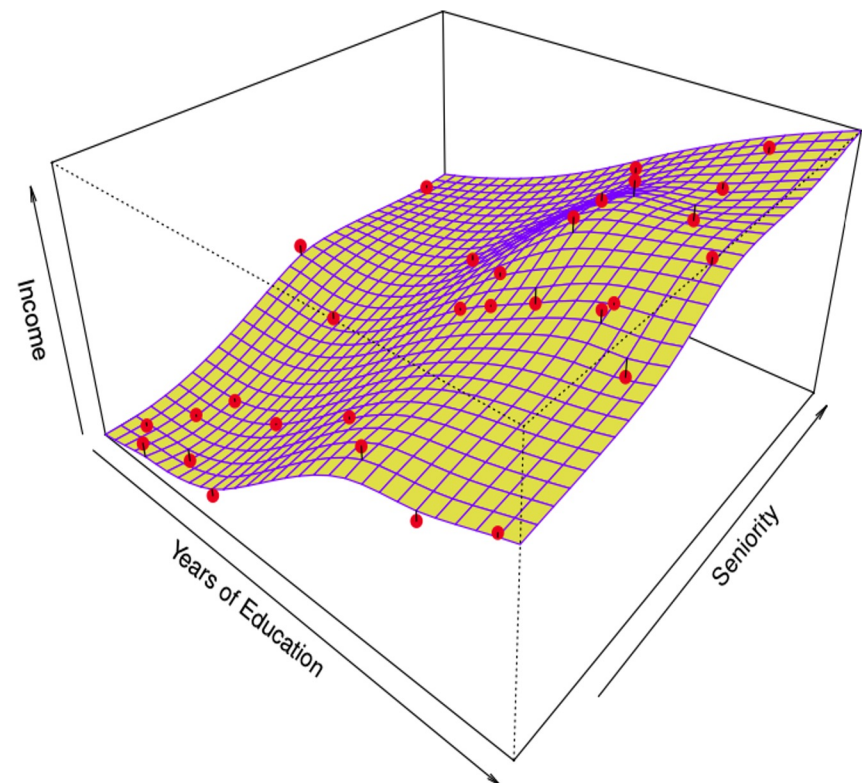


Sala í þúsundum stykkja sem fall af
auglýsingaútgjöldum í þremur
miðlum á 200 mismunandi
mörkuðum

Línulegt aðhvarf fyrir tekjugögnin:



Splæsifallaaðhvarf fyrir tekjugögnin



Túlkun og sveiginaleiki

(Hægt að fórna öðru fyrir hitt)



Breytur, úttak, líkön o.fl.

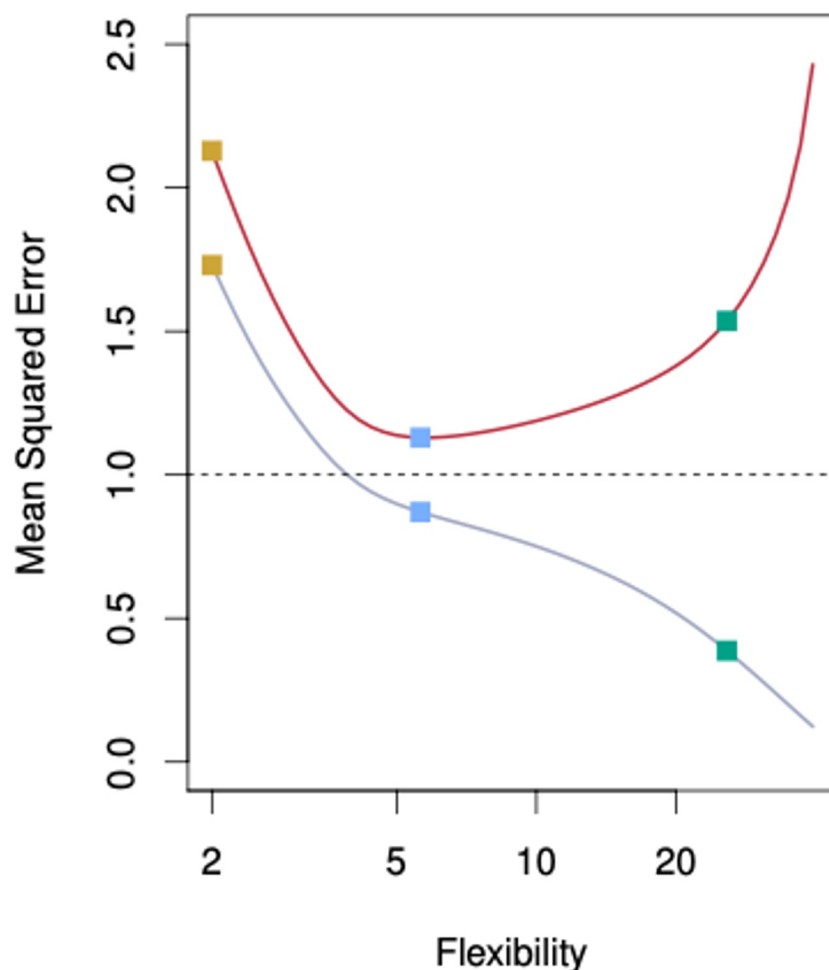
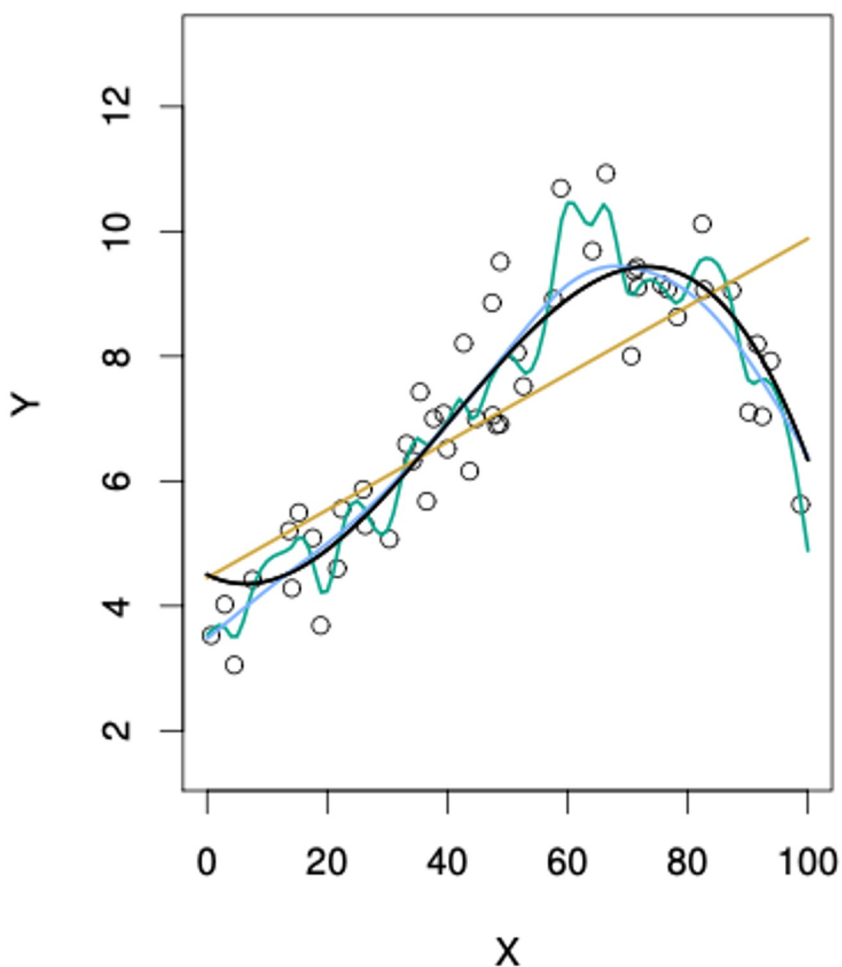
Líkan er ritað: $Y = f(X) + \epsilon$

Metið líkan og spá: $\hat{Y} = \hat{f}(X)$

Mælikvarði á gæði (MSE = mean squared error): $MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{f}(x_i))^2$

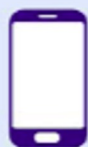
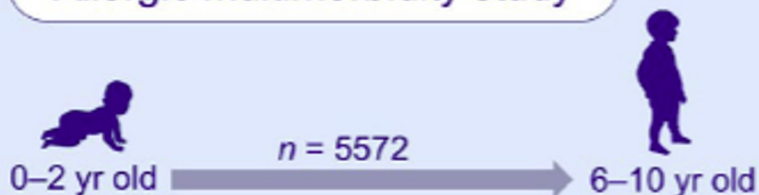
Þjálfun og test gögn, "Overfitting" o.fl.

Sjá bls. 31–32 í SL-bók



Myndir úr ofnæmisgrein

Allergic multimorbidity study



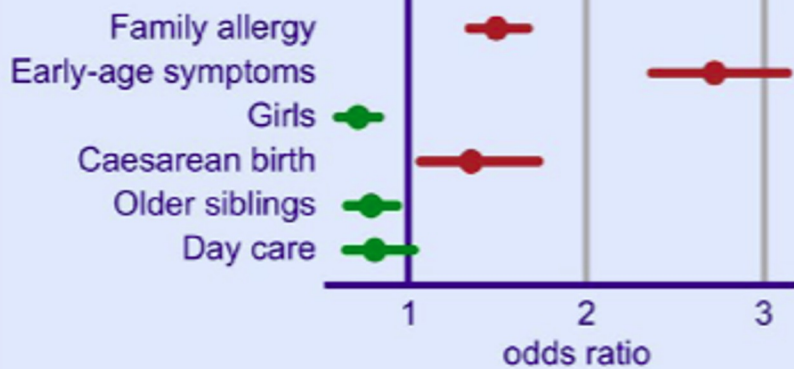
Phone interviews



Online questionnaire

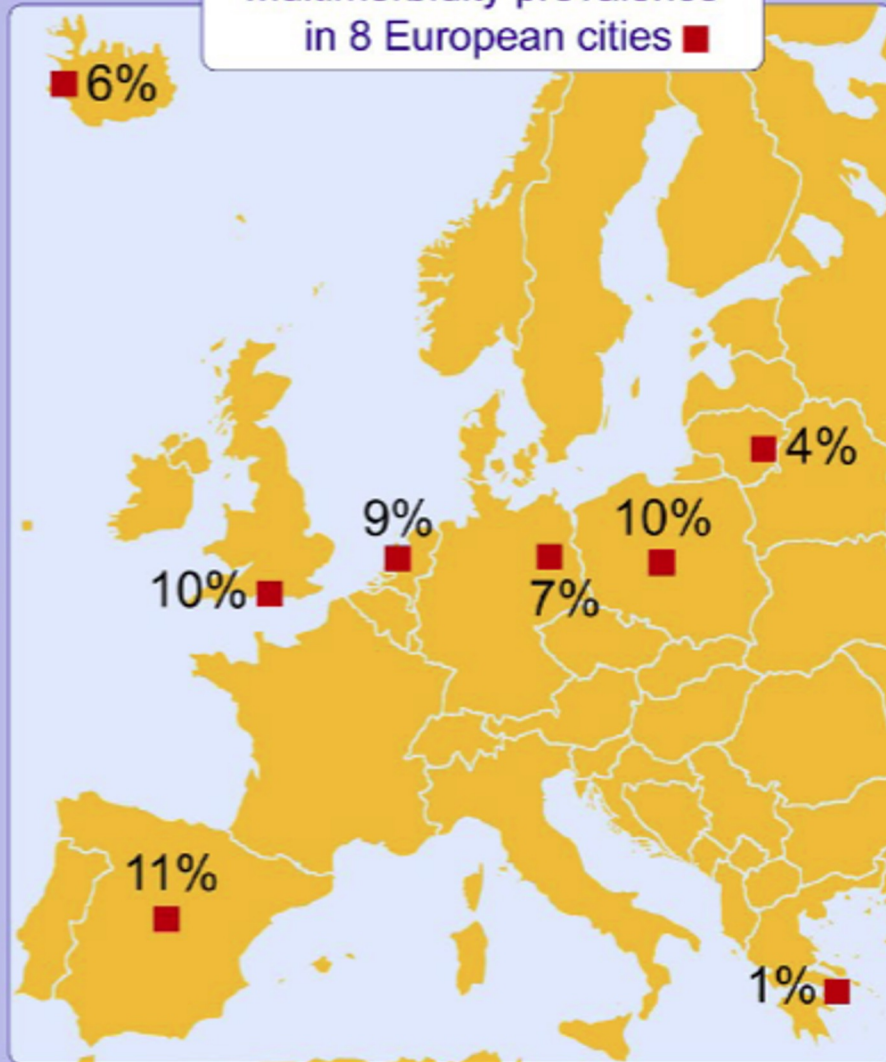
Allergic multimorbidity = two or all of at school age

Asthma
Rhinitis
Eczema



Protective and risk factors

Multimorbidity prevalence in 8 European cities



Í ofnæmisrannsókn var gerð tvíkostaaðhvarfsgreining (logistic regression) á fjölofnæmissjúkdómum í 6–10 ára börnum í 8 evrópskum borgum.

2.4 | Data processing and statistical analysis

Data were processed with Unix and MATLAB (version 9.3; The MathWorks Inc.). To assess the significance of covariates, a logistic regression model was constructed, where centre effect was included as a multiplicative factor:

Odds = (center effect) × (effect of covariate 1) × (effect of covariate 2) × ...

The covariates entering the model were selected using the AICc criterion (Akaike information criterion with correction). One advantage of using AICc is that it is independent of (arbitrary) statistical significance levels: it selects the model with the smallest expected mean squared error.¹³ Forest plots were used to show 95% confidence intervals.

Odds eða gagnlíkindi eru $\frac{p}{1-p}$ þar sem p = líkur á að fá sjúkdóm (ofnæmi).

Eftirfarandi mynd sýnir áhrifaþætti (effect of covariates) fyrir astma en á bls. á undan eru áhrifaþættir fjölofnæmis sýndir (sjá [Tölfræði frá grunni](#) k. 4.5.2).

