

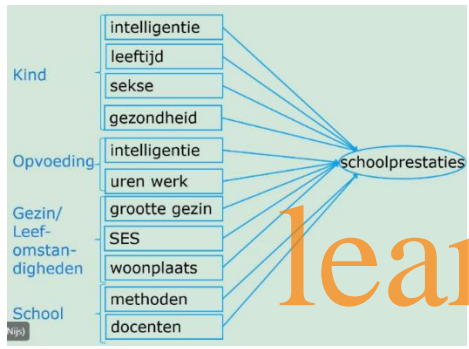
Hoorcollege 1 Meerdere regressie

Multipere regressie

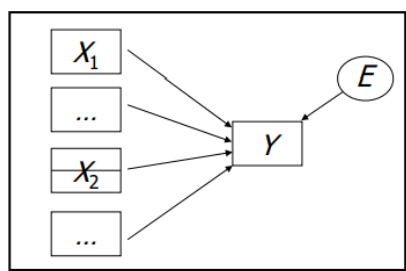
The image shows a page from a PDF document with several red boxes highlighting specific sections. The sections include 'Risicofactoren', 'Opvoedingsprincipes', 'Schoolfactoren', and 'Docenten'. The text is dense and appears to be a lecture note or a research paper. At the bottom of the page, there is a table of contents with the following entries: 'Risicofactoren', 'Opvoedingsprincipes', 'Schoolfactoren', 'Docenten', and 'Instituut'.

Voorbeeldverslag → allemaal factoren (risicofactoren, opvoedingsprincipes, schoolfactoren etc.) die een rol spelen op de schoolprestaties van kinderen. → nagaan wat de invloed van deze variabelen (onafhankelijk) hebben op de schoolprestaties (afhankelijke) → multipere regressie

Model



Padmodel multipere regressie



Met streepje: dichotome variabele met 2 categorieën

Geen streepje in: interval/ratio niveau

- 1 afhankelijke variabele (Y)
- 1 of meerdere onafhankelijke variabelen (minimaal interval)
- 1 of meerdere onafhankelijke variabele (dichotoom)

Onderwerpen college

- Regressiemodel
- Predictoren/onafhankelijke variabelen en afhankelijke variabelen
- (Gestandaardiseerde) regressiecoëfficiënten

- Kleinste kwadraten criterium
- Goodness-of-fit
- Toetsen van R^2
- Toetsen van B 's en β 's
- Vergelijking van ΔR^2
- Categorische kenmerken in regressiemodel; dummy's
- Assumpties toepassing in regressieanalyse

Voorbeeld

Onderzoeksvraag: kunnen we kennis van literatuur bij jongvolwassenen voorspellen met persoons-, gezins- en schoolkenmerken?

Populatie: jongvolwassenen

Variabelen: afhankelijke variabele y (kennis van literatuur), onafhankelijke variabele X (predictoren) (persoonlijke kenmerken, kenmerken ouderlijk huis, kenmerken school)

- ➔ Voor de populatie beschrijven en toetsen van de relaties tussen afhankelijke variabele Y en de predictoren X

Multipel regressie algemeen

Onderzoeksvraag: Kunnen we iemands waarde op een kenmerk voorspellen met kennis over andere kenmerken?

Doelen analyse:

- Beschrijven lineaire relaties tussen variabelen (regressiemodel)
- Toetsen hypothesen op relaties (significantie)
- Kwantificeren van relaties (effectgrootte)
- Kwalificeren van relaties (klein, middelmatig, groot)
- Beoordelen relevantie relaties (subjectief)
- Voorspellen van iemands waarde met regressiemodel (puntschatting en intervalschatting)

Waarschuwing: doe op basis van statistische samenhang GEEN uitspraken over causaliteit

Variabelen in voorbeeld

Read: kennis literatuur respondent (Y)

Fath_rd: kennis literatuur vader (X_1)

Moth_rd: kennis literatuur moeder (X_2)

Par_book: aantal boeken in ouderlijk huis (X_3)

Sch_rd: aandacht voor literatuur school (X_4)

Hist_rd: lezen verleden (X_5)

Educ: opleidingsniveau (X_6)

Meetniveau variabelen

Afhankelijke variabele Y :

- Gemeten op minimaal interval meetniveau

Meetniveau onafhankelijke variabelen X_k :

- Kenmerk gemeten op minimaal interval meetniveau

- Categorische kenmerk met 2 categorieën; nominaal meetniveau met 2 categorieën noemen we dichotoom
- Categorische kenmerk met meer dan 2 categorieën; nominaal/ordinaal meetniveau wordt omgezet in dummyvariabelen

Variabelen in voorbeeld

Allemaal minimaal interval meetniveau

Analyse techniek → multipele lineaire regressieanalyse (MR)

Regressiemodel (1)

Vergelijking Y

Voor geobserveerde variabele Y

Uitkomst (Y) = model (X) = voorspellingsfout

Voorspellingsfout = residual of error

Model = lineair regressiemodel

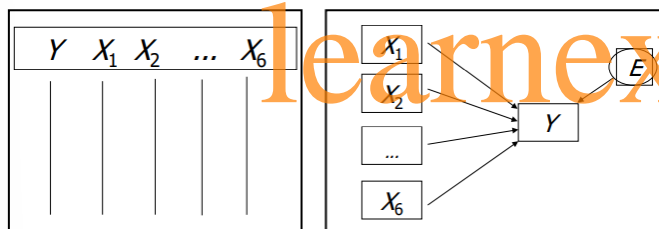
Vergelijking \hat{Y}

Voor voorspellen van waarde op \hat{Y} (dakje wijst op voorspelling)

Geschatte uitkomst (\hat{Y}) = model(X)

Verschil observatie en schatting = voorspellingsfout

Regressiemodel (2)



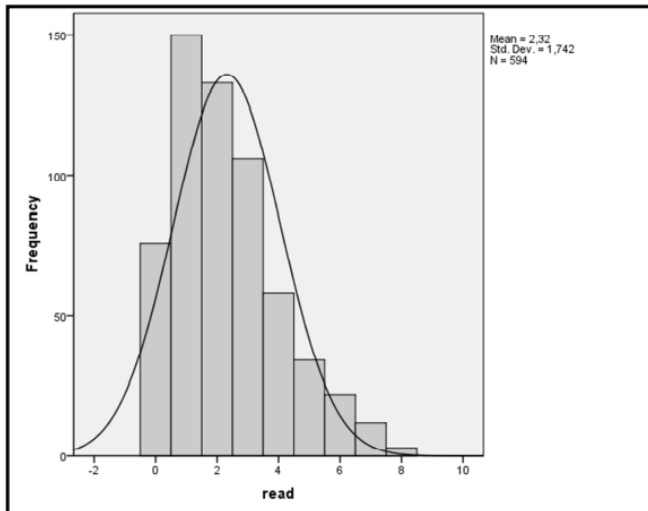
$$Y = B_0 + B_1X_1 + \dots + B_6X_6 + E$$

- Y = afhankelijke variabele (*dependent*)
- X = onafhankelijke variabelen (*predictors*)
- B_0 = intercept (*constant*), ook wel a
- B_1 = regressiecoëfficiënt (*slope*)
- E = voorspellingsfout (*error of residual*)

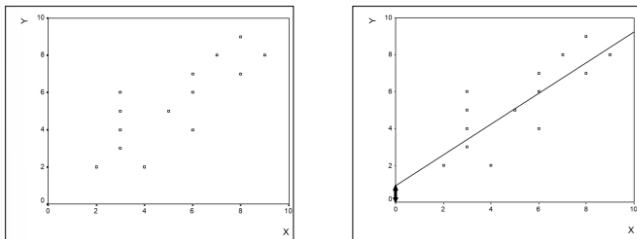
SPSS Datamatrix

| | read | fath_us | moth_us | par_book | sch_us | hst_us | educ | var | var | var | var |
|----|------|---------|---------|----------|--------|----------|------|-----|-----|-----|-----|
| 1 | 5 | 4.83 | 1.56 | 7.00 | 3.00 | 2.33 | 7 | | | | |
| 2 | 2 | 3.73 | 4.95 | 6.00 | 8.00 | 3.33 | 7 | | | | |
| 3 | 4 | 4.97 | 5.33 | 5.00 | 7.00 | 2.50 | 7 | | | | |
| 4 | 0 | 3.18 | 2.41 | 7.00 | 5.00 | 3.00 | 6 | | | | |
| 5 | 0 | 3.55 | 3.65 | 6.00 | 0 | 2.87 | 2 | | | | |
| 6 | 5 | 0 | 1.49 | 7.00 | 4.00 | 2.00 | 6 | | | | |
| 7 | 4 | 2.06 | 4.66 | 5.00 | 6.00 | 2.33 | 6 | | | | |
| 8 | 1 | 0 | 1.49 | 6.00 | 3.00 | 2.33 | 11 | | | | |
| 9 | 1 | 0 | 0 | 4.00 | 0 | 2.67 | 11 | | | | |
| 10 | 4 | 3.88 | 0 | 8.00 | 8.00 | 3.00 | 5 | | | | |
| 11 | 3 | 3.61 | 2.33 | 8.00 | 9.00 | 3.00 | 4 | | | | |
| 12 | 0 | 0 | 5.33 | 2.00 | 0 | 2.00 | 2 | | | | |
| 13 | 1 | 0 | 0 | 4.00 | 0 | 99999.00 | 2 | | | | |
| 14 | 0 | 0 | 1.14 | 5.00 | 0 | 3.00 | 11 | | | | |
| 15 | 0 | 0 | 3.43 | 6.00 | 0 | 2.33 | 2 | | | | |
| 16 | 1 | 0 | 4.61 | 5.00 | 0 | 2.67 | 2 | | | | |
| 17 | 1 | 0 | 0 | 1.00 | 0 | 3.00 | 3 | | | | |
| 18 | 2 | 0 | 1.49 | 1.00 | 0 | 4.00 | 11 | | | | |
| 19 | 0 | 0 | 0 | 0 | 0 | 99999.00 | 11 | | | | |
| 20 | 2 | 0 | 1.32 | 3.00 | 4.00 | 2.50 | 4 | | | | |
| 21 | 2 | 1.49 | 1.32 | 6.00 | 0 | 2.00 | 4 | | | | |
| 22 | 3 | 3.18 | 2.41 | 3.00 | 0 | 2.00 | 11 | | | | |
| 23 | 1 | 2.06 | 1.14 | 2.00 | 0 | 2.00 | 2 | | | | |
| 24 | 1 | 0 | 2.40 | 6.00 | 8.00 | 4.00 | 4 | | | | |
| 25 | 3 | 3.03 | 3.05 | 7.00 | 7.00 | 3.00 | 5 | | | | |
| 26 | 1 | 0 | 3.65 | 4.00 | 10.00 | 2.00 | 4 | | | | |
| 27 | 3 | 2.96 | 0 | 8.00 | 8.00 | 1.67 | 3 | | | | |

Histogram 'read'



Spreidingsdiagram



Vergelijking voor enkelvoudige regressie

- 1. Intercept of constante (b_0)
- 2. Regressiecoëfficiënt (b_1)

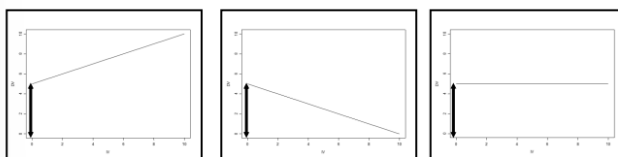
$$Y = b_0 + b_1 X_1$$

learnexams

Intercept = soort begin van de lijn, positie van Y waar X=0 geldt. Bij multiële regressie is t waar alle X=en 0 zijn.

Regressie is hoe stijl de lijn loopt.

Regressiecoëfficiënten



Intercept

$b_0 = 5$ $b_0 = 5$ $b_0 = 5$

Richting verband

$b_1 = 0.5$ $b_1 = -0.5$ $b_1 = 0$
 (positief) (negatief)

Interpretatie b_1 -waarde

De verandering in \hat{Y} bij toename van één eenheid in X .

Links = positieve relatie

Midden = negatieve relatie

Rechts = afwezigheid relatie