

Opgave 1.

Geef voor elk van de volgende modellen aan waarin ze verschillen van een lineair regressiemodel:

- a. variantie-analyse model,
- b. niet-lineair model,
- c. gegeneraliseerd lineair model,
- d. tijdreeksmodel.

Opgave 2.

Veronderstel dat aan je gevraagd wordt een tijdreeksanalyse uit te voeren op een gegeven data set $X_1, \dots, X_n$. Beschrijf kort, maar zo volledig mogelijk, de stappen die je gaat uitvoeren om aan deze opdracht te voldoen.

Opgave 3.

Zij

$$Y_i = f(x_i, \theta) + \varepsilon_i, \quad i = 1, \dots, n$$

een niet-lineair regressiemodel met $Y_1, \dots, Y_n$ onafhankelijke stochastische responsvariabelen, x_i een k -dimensionale vector van verklarende variabelen behorende bij Y_i ($i = 1, \dots, n$), ε_i een stochastische meetfout met verwachting 0 en variantie σ^2 behorende bij Y_i ($i = 1, \dots, n$), θ een onbekende p -dimensionale parametervector en f een niet-lineaire functie van θ .

- a. Hoe is de kleinste kwadratenschatter $\hat{\theta}$ van θ gedefinieerd?
- b. Zij $\widehat{\text{Cov}}(\hat{\theta}) = \hat{\sigma}^2(\hat{V}^T \hat{V})^{-1}$ een schatter voor de covariantiematrix van $\hat{\theta}$. Geef een, bij benadering, 95% betrouwbaarheidsinterval voor θ_l , $l = 1, \dots, p$ gebaseerd op de klassieke theorie.
- c. Formuleer de drie stappen van het bootstrapschema dat nodig is om een bootstrap betrouwbaarheidsinterval met bij benadering 95% betrouwbaarheid voor θ_l gegeven door

$$\left[2\hat{\theta} - \theta_{[(1-0.025)B]+1}^*, 2\hat{\theta} - \theta_{[0.025B]+1}^* \right] \quad (1)$$

te verkrijgen. (Je kunt gebruiken dat $\tilde{\varepsilon}_i$ het gecentreerde residu is, d.w.z. $\tilde{\varepsilon}_i = \hat{\varepsilon}_i - 1/n \sum_{i=1}^n \hat{\varepsilon}_i$ met $\hat{\varepsilon}_i = R_i(\hat{\theta}) = Y_i - f(x_i, \hat{\theta})$.)

d. Men heeft de beschikking over twee data sets waarvan men vermoedt dat hetzelfde niet-lineaire model van toepassing is. Het idee is dat de parameterwaarden voor de beide sets hetzelfde zijn, behalve die van θ_4 . Men bepaalt voor de beide data sets de twee soorten betrouwbaarheidsintervallen met bij benadering 95% betrouwbaarheid voor θ_4 . Wat zou je concluderen wanneer de klassieke betrouwbaarheidsintervallen voor de verschillende data sets elkaar gedeeltelijk overlappen, maar de bootstrap betrouwbaarheidsintervallen elkaar niet overlappen, terwijl de bootstrap betrouwbaarheidsintervallen voor beide sets aanzienlijk kleiner zijn dan de overeenkomstige klassieke betrouwbaarheidsintervallen?

Opgave 4.

Gegeven zijn de waarden van n (n groot) realisaties van een responsvariabele met mogelijke waarden 0 en 1. Men wenst het verband te onderzoeken tussen deze responsvariabele en vier andere variabelen, te weten twee continue variabelen gemeten op een ratio schaal en twee variabelen gemeten op een nominale schaal met, respectievelijk, twee en vier categorieën.

- a. Stel een logistisch regressiemodel op voor dit probleem.
- b. Hoe kan men per variabele toetsen of deze variabele in het model opgenomen dient te worden?

Opgave 5.

Hieronder worden enkele praktijksituaties geschetst. Kies voor elke situatie één of meer geschikte modellen. Kies hierbij uit: lineair regressiemodel, variantie-analyse model, niet-lineair model, gegeneraliseerd lineair model, tijdreeksmodel. Beargumenteer je antwoord. Beschrijf het gekozen model zo volledig mogelijk. Vermeld hierbij in elk geval wat de respons- en verklarende variabelen zijn.

- a. Een verzekeringsmaatschappij wil voor verschillende vliegtuigtypen onderzoeken of het aantal schadeclaims afhangt van variabelen als bouwjaar, totale onderhoudstijd en het aantal vliegers.
- b. In een bepaald industrieel proces komen onregelmatig kleine fouten voor. Om de oorzaak te onderzoeken, wordt het volgende experiment gedaan: verschillende partijen ruw materiaal worden geselecteerd en elke partij wordt in twee gedeeltes gesplitst. Voor elke partij wordt het ene deel met de standaard methode behandeld, het andere deel met een gemodificeerde behandeling waarin de temperatuur gedurende een tijdje wordt verlaagd. Vóór de behandeling wordt aan elke partij in z'n geheel de zuiverheidsgraad van het materiaal gemeten. Tijdens de behandelingen wordt genoteerd of ja dan nee de kleine fouten optreden.
- c. Om te onderzoeken om de hoeveel tijd en hoeveel stuks van een bepaald produkt dient te worden bijbesteld, registreren verschillende vestigingen van een grote supermarktketen gedurende een jaar hoeveel er elke dag waarop de vestiging open is, van

het produkt wordt verkocht. Er wordt een model opgesteld dat dient ter analyse van de gegevens van één vestiging.

Normering:

1.a: 2	2: 6	3.a: 1	4.a: 3	5.a: 3
1.b: 2		3.b: 1	4.b: 3	5.b: 3
1.c: 2		3.c: 2		5.c: 3
1.d: 2		3.d: 3		

$$\text{Eindcijfer tentamen} = \text{totaal}/4 + 1$$