

Aan: Bestuur van de Vereniging VSE

Van: GKv Amersfoort-Zuid

Datum: 3 april 2023

Status: Definitief

Onderwerp: Amendement o.b.v artikel 21 lid 2 van de statuten ten behoeve van algemene vergadering van 12 april 2023

1. Inleiding

In het kader van de besluitvorming staat de volgende tekst:

De voorlopige schuld per 2027 wordt bepaald op basis van (3B3):

- - De door Triple A opgestelde berekeningen ([voor rapport, klik hier](#)).
- - De vermogenspositie per 31-12-2022.
- - Verwacht rendement van 4% (zie voor onderbouwing van aanpassing van 3,5% naar 4% de bovengenoemde notitie bij berekeningen Triple A).
- ***Het gemiddeld zielental per kerk over de jaren 2020, 2021 en 2022. Voor de definitieve bepaling van de schuld in 2026 wordt het ledental bepaald door de ontwikkeling van ledentallen van de jaren 2018 t/m 2022 te extrapoleren naar 2027. Dit om tegemoet te komen aan het verzoek van kerken om rekening te houden met de reeds ingezette daling van het aantal leden. Bij samenwerkingsgemeenten wordt alleen het GKv-deel geteld.***

“De voorlopige schuld per 2027 (in 2024 aan de kerken te sturen) bedraagt € 441,79/ziel (3B4).”

Extrapoleren kan op vele manieren geschieden. Exponentieel, lineair en nog een aantal andere varianten. Verschillende berekeningen kunnen tot grote verschillen leiden. In de bijlagen zijn voor een paar voorbeeldkerken berekeningen gemaakt. Voor ieder van de kerken komt de eindnota op hetzelfde bedrag uit als alleen 2020, 2021 en 2022 gemiddeld worden. Bij extrapoleren kan het alle kanten opvliegen. Een kerk heeft waarschijnlijk grote ‘pech’ als er in 2022 een paar gezinnen zijn bijgekomen en ‘geluk’ als er relatief veel leden vertrokken zijn in 2018 en/of 2019. Maar waarom zou dat doorzetten?

In het voorbeeld is eenvoudig lineair geëxtrapoleerd door in excel de jaren 2018 tot en met 2022 te selecteren en vervolgens door naar beneden te slepen de andere jaren gevuld te krijgen.

Wij kunnen ons niet vinden in het besluit om te gaan extrapoleren en wel om de volgende redenen.

- Onvoldoende wordt duidelijk gemaakt wat het probleem precies is;
- Niet duidelijk wordt gemaakt dat extrapoleren dan de oplossing voor het probleem is;
- Niet duidelijk wordt gemaakt aan welke wijze van extrapoleren gedacht wordt en dat het verantwoord is om die methode te kiezen;
- Wij betwijfelen of extrapoleren leidt tot “faire” uitkomsten. Zie bijlage 1 voor enkele voorbeelden;
- Waarom geoordeeld is dat op basis van slechts 5 datapunten, de ledenaantallen van 2018 tot en met 2022, het verantwoord is om te extrapoleren? (Wij kunnen ons niet voorstellen dat of de actuaaris in het bestuur of de extern actuaaris van de VSE op schrift wil stellen dat een en ander verdedigbaar is). Zie ook bijlage 2 en bijlage 3 (vraag voor de aardigheid en ter illustratie ook voorgelegd aan Chat GPT).

2. Amendement

Wij stellen voor (**amendement**) uit de voorgelegde teksten de volgende passage te schrappen:

“Voor de definitieve bepaling van de schuld in 2026 wordt het ledental bepaald door de ontwikkeling van ledentallen van de jaren 2018 t/m 2022 te extrapoleren naar 2027. Dit om tegemoet te komen aan het verzoek van kerken om rekening te houden met de reeds ingezette daling van het aantal leden. Bij samenwerkingsgemeenten wordt alleen het GKv-deel geteld.”

3. Afsluitende opmerking

Als op een later moment ons duidelijk kan worden gemaakt wat precies het probleem is dat moet worden opgelost, dat extrapoleren de juiste methode is om het probleem op te lossen en dat extrapoleren leidt tot faire en te verantwoorden (gelet op aantal datapunten) uitkomsten, dan willen wij graag op een later tijdstip voorstellen in overweging nemen. In het kader van de transparantie en om de voorstellen beter te kunnen beoordelen zien wij dan ook graag een lijst tegemoet van de aantallen leden per kerk berekend op basis van het driejaars gemiddelde vergeleken met de aantallen berekend met behulp van extrapolatie.

Wat ons betreft is het nu nog te vroeg om te besluiten al dan niet te extrapoleren. Zou waarschijnlijk wel in ons “voordeel” zijn, maar of het ook “fair” is betwijfelen wij vooralsnog.

Graag tot nadere toelichting bereid,

J. Kool, o.b.v. separaat toegemailde machtiging



Bijlage 1

Leden		Kerk A	Kerk B	Kerk C	Kerk D	Kerk E	Kerk F	
								Totaal
2018	1	150	330	270	300	375	330	1755
2019	2	200	315	270	300	345	315	1745
2020	3	240	285	295	295	315	275	1705
2021	4	295	285	285	285	285	285	1720
2022	5	320	285	275	275	255	295	1705
Gemiddelde 2020 t/m 2022		285	285	285	285	285	285	1710
Procentueel aandeel in tekort		16,67%	16,67%	16,67%	16,67%	16,67%	16,67%	100,00%
Eindnota 441,79 per ziel		125.910	125.910	125.910	125.910	125.910	125.910	755.461
Extrapolatie naar 2027								
2023	6	371,5	264	286,5	271,5	225	270	1689
2024	7	415	252	289	265	195	260	1676
2025	8	458,5	240	291,5	258,5	165	250	1664
2026	9	502	228	294	252	135	240	1651
2027	10	545,5	216	296,5	245,5	105	230	1639
		33,29%	13,18%	18,10%	14,98%	6,41%	14,04%	100,00%
Eindnota		251.513	99.591	136.707	113.192	48.412	106.046	755.461
Verschil		125.603	-26.319	10.797	-12.718	-77.498	-19.864	
Kerk A		Jonge kerk in nieuwbouwwijk						
Kerk B		Daling in 2018 en 2019, nu stabiel						
Kerk C		Redelijk stabiel, laatste jaren lichte daling na forse stijging in 2019						
Kerk D		Laatste jaren conform kerk C, maar in 2018 en 2019 meer leden dan kerk C						
Kerk E		Kerk met sterke daling zielen, maar grote reserves in verleden opgebouwd						
Kerk F		Weer groeiende kerk door nieuwe predikant, weer op weg naar niveau van 2019						

One in ten rule

[Add languages](#)
[Article](#) [Talk](#)
[Read](#) [Edit](#) [View history](#) [Tools](#)

From Wikipedia, the free encyclopedia

In *statistics*, the **one in ten rule** is a *rule of thumb* for how many *predictor* parameters can be estimated from data when doing *regression analysis* (in particular *proportional hazards models* in *survival analysis* and *logistic regression*) while keeping the risk of *overfitting* low. The rule states that one predictive variable can be studied for every ten events.^{[1][2][3][4]} For logistic regression the number of events is given by the size of the smallest of the outcome categories, and for survival analysis it is given by the number of *uncensored* events.^[3]

For example, if a sample of 200 patients is studied and 20 patients die during the study (so that 180 patients survive), the one in ten rule implies that two pre-specified predictors can reliably be fitted to the total data. Similarly, if 100 patients die during the study (so that 100 patients survive), ten pre-specified predictors can be fitted reliably. If more are fitted, the rule implies that overfitting is likely and the results will not predict well outside the *training data*. It is not uncommon to see the 1:10 rule violated in fields with many variables (e.g. gene expression studies in cancer), decreasing the confidence in reported findings.^[5]

Improvements [edit]

A "one in 20 rule" has been suggested, indicating the need for *shrinkage* of regression coefficients, and a "one in 50 rule" for *stepwise selection* with the default *p-value* of 5%.^{[4][6]} Other studies, however, show that the one in ten rule may be too conservative as a general recommendation and that five to nine events per predictor can be enough, depending on the research question.^[7]

More recently, a study has shown that the ratio of events per predictive variable is not a reliable statistic for estimating the minimum number of events for estimating a logistic prediction model.^[8] Instead, the number of predictor variables, the total sample size (events + non-events) and the events fraction (events / total sample size) can be used to calculate the expected prediction error of the model that is to be developed.^[9] One can then estimate the required sample size to achieve an expected prediction error that is smaller than a predetermined allowable prediction error value.^[9]

Alternatively, three requirements for prediction model estimation have been suggested: the model should have a global shrinkage factor of $\geq .9$, an absolute difference of $\leq .05$ in the model's apparent and adjusted *Nagelkerke R²*, and a precise estimation of the overall risk or rate in the target population.^[10] The necessary sample size and number of events for model development are then given by the values that meet these requirements.^[10]

References [edit]

- ↑ Harrell, F. E. Jr.; Lee, K. L.; Califf, R. M.; Pryor, D. B.; Rosati, R. A. (1984). "Regression modelling strategies for improved prognostic prediction". *Stat Med*. **3** (2): 143–52. doi:10.1002/sim.4780030207. PMID 6463451.
- ↑ Harrell, F. E. Jr.; Lee, K. L.; Mark, D. B. (1996). "Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors" (PDF). *Stat Med*. **15** (4): 361–87. doi:10.1002/(sici)1097-0258(19960229)15:4<361::aid-sim168>3.0.co;2-4. PMID 8668867.
- ↑ ^a ^b Peduzzi, Peter; Concato, John; Kemper, Elizabeth; Holford, Theodore R.; Feinstein, Alvan R. (1996). "A simulation study of the number of events per variable in logistic regression analysis". *Journal of Clinical Epidemiology*. **49** (12): 1373–1379. doi:10.1016/s0895-4356(96)00236-3. PMID 8970487.
- ↑ ^a ^b "Chapter 8: Statistical Models for Prognostication: Problems with Regression Models". Archived from the original on October 31, 2004. Retrieved 2013-10-11.
- ↑ Ernest S. Shtatland, Ken Kleinman, Emily M. Cain. Model building in Proc PHREG with automatic variable selection and information criteria. Paper 206–30 in SUGI 30 Proceedings, Philadelphia, Pennsylvania April 10–13, 2005. <http://www2.sas.com/proceedings/sugi30/206-30.pdf>
- ↑ Steyerberg, E. W.; Eijkemans, M. J.; Harrell, F. E. Jr.; Habbema, J. D. (2000). "Prognostic modelling with logistic regression analysis: a comparison of selection and estimation methods in small data sets". *Stat Med*. **19** (8): 1059–1079. doi:10.1002/(sici)1097-0258(20000430)19:8<1059::aid-sim412>3.0.co;2-0. PMID 10790680.
- ↑ Vittinghoff, E.; McCulloch, C. E. (2007). "Relaxing the Rule of Ten Events per Variable in Logistic and Cox Regression". *American Journal of Epidemiology*. **165** (6): 710–718. doi:10.1093/aje/kwk052. PMID 17182981.
- ↑ van Smeden, Maarten; de Groot, Joris A. H.; Moons, Karel G. M.; Collins, Gary S.; Altman, Douglas G.; Eijkemans, Marinus J. C.; Reitsma, Johannes B. (2016-11-24). "No rationale for 1 variable per 10 events criterion for binary logistic regression analysis". *BMC Medical Research Methodology*. **16** (1): 163. doi:10.1186/s12874-016-0267-3. ISSN 1471-2288. PMC 5122171. PMID 27881078.
- ↑ ^a ^b van Smeden, Maarten; Moons, Karel Gm; de Groot, Joris Ah; Collins, Gary S.; Altman, Douglas G.; Eijkemans, Marinus Jc; Reitsma, Johannes B. (2018-01-01). "Sample size for binary logistic prediction models: Beyond events per variable criteria". *Statistical Methods in Medical Research*. **28** (8): 2455–2474. doi:10.1177/0962280218784726. ISSN 1477-0334. PMC 6710621. PMID 29966490.
- ↑ ^a ^b Riley, Richard D.; Snell, Kym IE; Ensor, Joie; Burke, Danielle L.; Jr, Frank E. Harrell; Moons, Karel GM; Collins, Gary S. (2018). "Minimum sample size for developing a multivariable prediction model: PART II - binary and time-to-event outcomes". *Statistics in Medicine*. **38** (7): 1276–1296. doi:10.1002/sim.7992. ISSN 1097-0258. PMC 6519266. PMID 30357870.

Bijlage 3

Model: GPT-4

SYSTEM MESSAGE: Je bent een derdejaars student statistiek

Als derdejaars student statistiek heb je al veel kennis opgedaan over de basisprincipes en technieken die in de statistiek worden gebruikt. Je hebt geleerd over kansrekening, beschrijvende statistiek, inferentiële statistiek, regressieanalyse en waarschijnlijk nog veel meer. Nu je in je derde jaar bevindt, is het tijd om dieper in te gaan op geavanceerdere onderwerpen en je vaardigheden verder te ontwikkelen. Je kunt je bijvoorbeeld verdiepen in multivariate analyse, tijdreeksanalyse, machine learning, Bayesian statistiek en experimenteel ontwerp.

Als je vragen hebt over statistiek of hulp nodig hebt bij een specifiek onderwerp, aarzel dan niet om te vragen. Ik ben hier om je te helpen!

Geef me voorbeelden met voorbeelddata waarom lineair interpoleren met 5 datapunten slecht is

Lineaire interpolatie is een methode om de waarde van een functie op een bepaald punt te schatten op basis van de bekende waarden van de functie op andere punten. Het werkt goed wanneer de onderliggende relatie tussen de variabelen lineair is of bijna lineair. Echter, bij niet-lineaire relaties kunnen de schattingen door lineaire interpolatie sterk afwijken van de werkelijke waarden.

Laten we een voorbeeld bekijken met een kwadratische functie (een niet-lineaire relatie) en vijf

datapunten: $x: [1, 2, 3, 4, 5]$ $y: [1, 4, 9, 16, 25]$

De onderliggende functie is $y = x^2$. We willen nu de waarde van y schatten bij $x = 2,5$ met behulp van lineaire interpolatie.

Als we kijken naar de punten $(2, 4)$ en $(3, 9)$, dan kunnen we de helling tussen deze twee punten berekenen: $(9 - 4) / (3 - 2) = 5$. Nu kunnen we de waarde van y schatten bij $x = 2,5$: $y = 4 + (2,5 - 2) * 5 = 6,5$.

De werkelijke waarde van y bij $x = 2,5$ is echter $(2,5)^2 = 6,25$. De geschatte waarde is dus $6,5$, wat een fout van $0,25$ oplevert.

In dit voorbeeld is de fout misschien niet al te groot, maar het probleem wordt duidelijker bij complexere of sterk niet-lineaire functies. Lineaire interpolatie gaat ervan uit dat de relatie tussen de variabelen constant is tussen de bekende punten, wat vaak niet het geval is.

Als je slechts vijf datapunten hebt, kan het moeilijk zijn om de onderliggende relatie tussen de variabelen te achterhalen. Lineaire interpolatie kan dan leiden tot onnauwkeurige schattingen. In dergelijke situaties zijn er andere, geavanceerdere interpolatiemethoden, zoals spline-interpolatie of polynomiale interpolatie, die betere schattingen kunnen opleveren bij niet-lineaire relaties.

Kan je me een datavoorbeeld geven waarom lineaire interpolatie om demografie te voorspellen een slecht idee is?

Stel dat we de bevolkingsgroei in een bepaald land willen voorspellen met behulp van lineaire interpolatie. De bevolking is afhankelijk van verschillende factoren, zoals geboortecijfers,

sterftcijfers en migratie. Al deze factoren kunnen in de loop van de tijd veranderen en hebben een niet-lineair effect op de bevolking.

Jaar: [2000, 2005, 2010, 2015, 2020] Bevolking: [10.000, 11.500, 13.000, 14.000, 14.800] (x 1.000)

We willen nu de bevolking voorspellen voor het jaar 2025 met behulp van lineaire interpolatie. We kijken naar de gegevens voor 2015 en 2020 en berekenen de helling: $(14.800 - 14.000) / (2020 - 2015) = 800 / 5 = 160$.

Vervolgens schatten we de bevolking in 2025: $14.800 + (2025 - 2020) * 160 = 14.800 + 5 * 160 = 15.600$.

Dit zou betekenen dat de bevolking in 2025 15,6 miljoen mensen zou bedragen. Echter, de werkelijke bevolking kan sterk afwijken van deze schatting, omdat er geen rekening is gehouden met de niet-lineaire effecten van geboortecijfers, sterftcijfers en migratie.

Bijvoorbeeld, als het geboortecijfer plotseling daalt of het sterftcijfer stijgt, kan de bevolking minder snel groeien dan de lineaire interpolatie suggereert. Aan de andere kant, als het land een instroom van immigranten ziet of een plotselinge stijging van het geboortecijfer, kan de bevolking veel sneller groeien dan voorspeld.

Lineaire interpolatie houdt geen rekening met deze niet-lineaire veranderingen en kan dus leiden tot onnauwkeurige voorspellingen van de bevolkingsgroei. Voor demografische voorspellingen zijn meer geavanceerde modellen, zoals cohort-component methoden, beter geschikt, omdat ze rekening houden met de afzonderlijke factoren die van invloed zijn op de bevolkingsgroei en de onderliggende niet-lineaire relaties.

Verander in het vorige antwoord de term bevolking voor kerkgangers, zorg dat het correct Nederlands blijft

Stel dat we het aantal kerkgangers in een bepaald land willen voorspellen met behulp van lineaire interpolatie. Het aantal kerkgangers is afhankelijk van verschillende factoren, zoals religieuze overtuigingen, demografische veranderingen en sociale trends. Al deze factoren kunnen in de loop van de tijd veranderen en hebben een niet-lineair effect op het aantal kerkgangers.

Laten we een hypothetisch voorbeeld bekijken met de volgende gegevens over kerkgangers (in duizenden): Jaar: [2000, 2005, 2010, 2015, 2020] Kerkgangers: [5.000, 4.500, 4.000, 3.500, 3.000]

We willen nu het aantal kerkgangers voorspellen voor het jaar 2025 met behulp van lineaire interpolatie. We kijken naar de gegevens voor 2015 en 2020 en berekenen de helling: $(3.000 - 3.500) / (2020 - 2015) = -500 / 5 = -100$.

Vervolgens schatten we het aantal kerkgangers in 2025: $3.000 + (2025 - 2020) * -100 = 3.000 + 5 * -100 = 2.500$.

Dit zou betekenen dat het aantal kerkgangers in 2025 2,5 miljoen mensen zou bedragen. Echter, het werkelijke aantal kerkgangers kan sterk afwijken van deze schatting, omdat er geen rekening is gehouden met de niet-lineaire effecten van religieuze overtuigingen, demografische veranderingen en sociale trends.

Bijvoorbeeld, als er een plotselinge toename van religieuze betrokkenheid is of een groeiende interesse in het bijwonen van kerkdiensten, kan het aantal kerkgangers veel sneller toenemen dan de lineaire interpolatie suggereert. Aan de andere kant, als de samenleving steeds meer seculariseert of als andere factoren het kerkbezoek ontmoedigen, kan het aantal kerkgangers sneller afnemen dan voorspeld.

Lineaire interpolatie houdt geen rekening met deze niet-lineaire veranderingen en kan dus leiden tot onnauwkeurige voorspellingen van het aantal kerkgangers. Voor een betere voorspelling van het aantal kerkgangers zijn meer geavanceerde modellen nodig, die rekening houden met de afzonderlijke factoren die van invloed zijn op kerkbezoek en de onderliggende niet-lineaire relaties.