

Algemene informatie voor vraag 1 t/m 3

Bij een enquête die met de computer wordt afgenomen (denk b.v. aan een internetenquête) is het vrij eenvoudig om niet alleen de antwoorden van een respondent te verzamelen, maar ook de responstijden. De responstijd, de tijd die een respondent er over doet om zijn antwoord te geven, kan geanalyseerd worden om extra informatie te verzamelen. Een korte responstijd kan aangeven dat de respondent over deze kwestie overtuigd is van zijn of haar mening, terwijl bij een langere responstijd de respondent minder zeker van zijn mening is.

Aan de hand van de responstijd op bijvoorbeeld de vraag “Het openbaar vervoer is een bruikbaar alternatief voor de auto”, is het mogelijk te zien in hoeverre iemand de kwestie echt overweegt. Hoe deze relatie precies ligt is echter nogal ingewikkeld en is ook niet het onderwerp van deze vraag. Het onderzoek dat in vraag 1 t/m 3 wordt beschreven, heeft als doel uit de responstijd de nadenkijd af te leiden.

In de responstijd zit, naast nadenkijd, ook de tijd die het kost om een vraag te lezen en te begrijpen. Er kunnen allerlei factoren van invloed zijn op deze leestijd die niets met de mening van de respondent te maken hebben. Denk bijvoorbeeld aan leeftijd, opleiding, geslacht, of de kennis van het Nederlands als eerste of tweede taal. Om een beeld te krijgen van de mogelijke invloeden van deze factoren wordt een experiment uitgevoerd, waaraan 100 vrijwilligers meedoen.

Op een computerscherm wordt een plaatje geprojecteerd met onder dat plaatje een vraag die over het plaatje gaat. Door de vraag langer of korter te maken kan de leestijd gemanipuleerd worden en door de vraag (of het plaatjes) ingewikkelder te maken kan de nadenkijd gemanipuleerd worden. Zo kan er een plaatje staan van een boom met aan de linkerkant een rode bal en aan de rechterkant een groen blok. De vraag kan dan luiden: “De kleur van de bal is” met als antwoordmogelijkheden “rood” en “groen”, of de vraag luidt: “U ziet hier een plaatje van een boom met aan een kant een bal en aan de andere kant een blok. De kleur van de bal is:” met dezelfde antwoordmogelijkheden. Beide vragen worden geacht even moeilijk te zijn, maar de eerste is veel korter dan de andere. Er kan ook een plaatje komen met dezelfde boom, maar naast de bal en het blok nog veel meer verschillende figuren, vergezeld van dezelfde vragen als voorgaand voorbeeld. De leestijden van beide vragen worden dan geacht om hetzelfde te zijn al in de eerste situatie, maar de vraag is ingewikkelder en de nadenkijd zou dus langer moeten zijn. Iedere deelnemer aan het experiment krijgt

zo vier vragen te beantwoorden: één makkelijke korte, één makkelijke lange, één moeilijke korte en één moeilijke lange.

Vraag 1

Als eerste besluit een onderzoekster om de responsietijden op de lange makkelijke vragen te onderzoeken. Het idee is dat hier zo weinig nadenktijd is dat verschillen in de responsietijd een direct gevolg van verschillen in de leestijd zijn. Zij gebruikt een lineair regressiemodel om te onderzoeken of de variabelen leeftijd (LEEFTIJD), geslacht (GESLACHT, gecodeerd als '0' voor de mannen en '1' voor de vrouwen), hoogste afgemaakte opleiding (OPLEID: 1: universiteit, HBO, MBO; 2: middelbare school, VMBO; 3 lagere school; 4: geen) en of het Nederlands de eerste taal is (NATIVE: '0' is 'nee' en '1' is 'ja') invloed hebben op de responstijd in seconden (RESPONS). Hieronder staan delen van de output.

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.881 ^a	.777	.768	2.78053

a. Predictors: (Constant), NATIVE, GESLACHT, OPLEID, LEEFTIJD

ANOVA^b

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	2558.320	4	639.580	82.725	.000 ^a
	Residual	734.481	95	7.731		
	Total	3292.801	99			

a. Predictors: (Constant), NATIVE, GESLACHT, OPLEID, LEEFTIJD

b. Dependent Variable: RESPONS

Coefficients^a

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	16.349	1.320		12.389	.000
	LEEFTIJD	.116	.037	.160	3.142	.002
	GESLACHT	-8.545	.566	-.744	-15.097	.000
	OPLEID	2.637	.315	.417	8.376	.000
	NATIVE	-6.138	.792	-.392	-7.750	.000

a. Dependent Variable: RESPONS

- a. (5 punten) Wat is de regressievergelijking voor de voorspelling van RESPONS. Neem alleen significante prediktoren in je vergelijking op (gebruik als α 0.05).

- b. (5 punten) Hoeveel wordt de op basis van dit model voorspelde RESPONS groter/kleiner wanneer iemand 10 jaar ouder wordt

Omdat de onderzoekster vermoedde dat het effect van leeftijd wel eens af zou kunnen hangen van het geslacht (immers vrouwen hebben nog steeds een langere levensverwachting, dus wellicht is de slijtage ook wel minder...) nam ze in een tweede blok een interactievariabele op. Deze variabele was het product van leeftijd en geslacht (LEEFGESL). Hieronder staan delen van de nieuwe output:

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Change Statistics				
					R Square Change	F Change	df1	df2	Sig. F Change
1	.881 ^a	.777	.768	2.78053	.777	82.725	4	95	.000
2	.900 ^b	.809	.799	2.58334	.033	16.057	1	94	.000

a. Predictors: (Constant), NATIVE, GESLACHT, OPLEID, LEEFTIJD

b. Predictors: (Constant), NATIVE, GESLACHT, OPLEID, LEEFTIJD, LEEFGESL

ANOVA^c

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	2558.320	4	639.580	82.725	.000 ^a
	Residual	734.481	95	7.731		
	Total	3292.801	99			
2	Regression	2665.477	5	533.095	79.881	.000 ^b
	Residual	627.324	94	6.674		
	Total	3292.801	99			

a. Predictors: (Constant), NATIVE, GESLACHT, OPLEID, LEEFTIJD

b. Predictors: (Constant), NATIVE, GESLACHT, OPLEID, LEEFTIJD, LEEFGESL

c. Dependent Variable: RESPONS

Coefficients^a

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	16.349	1.320		12.389	.000
	LEEFTIJD	.116	.037	.160	3.142	.002
	GESLACHT	-8.545	.566	-.744	-15.097	.000
	OPLEID	2.637	.315	.417	8.376	.000
	NATIVE	-6.138	.792	-.392	-7.750	.000
2	(Constant)	12.147	1.613		7.529	.000
	LEEFTIJD	.239	.046	.332	5.196	.000
	GESLACHT	.432	2.301	.038	.188	.851
	OPLEID	2.687	.293	.425	9.180	.000
	NATIVE	-6.010	.737	-.384	-8.160	.000
	LEEFGESL	-.265	.066	-.850	-4.007	.000

a. Dependent Variable: RESPONS

- c. (5 punten) Waaraan kun je zien dat het de moeite waard is om de interactievariabele (LEEFGESL) in het model op te nemen?
- d. (5 punten) Hoeveel wordt de op basis van dit laatste model (dus met LEEFGESL) voorspelde RESPONS groter/kleiner wanneer een man 10 jaar ouder wordt en hoeveel wanneer een vrouw 10 jaar ouder wordt?
- e. (5 punten) De variabele OPLEID is als een continue variabele in het model opgenomen. Wat impliceert dat voor de interpretatie van het regressiegewicht? Is deze aanname reëel?

De variabele OPLEID is gehercodeerd in 3 dummyvariabelen: OPLEID1 is 1 als OPLEID gelijk is aan 1 (en anders is OPLEID1 gelijk aan 0), OPLEID2 is 1 als OPLEID gelijk is aan 2 en OPLEID3 is 1 als OPLEID gelijk is aan 3. Hieronder staan delen van de output wanneer deze drie dummyvariabelen in plaats van de variabele OPLEID in het model worden opgenomen:

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.944 ^a	.892	.883	1.97026

a. Predictors: (Constant), OPLEID3, GESLACHT, NATIVE, LEEFTIJD, OPLEID2, OPLEID1, LEEFGESL

ANOVA^b

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	2935.666	7	419.381	108.035	.000 ^a
	Residual	357.135	92	3.882		
	Total	3292.801	99			

a. Predictors: (Constant), OPLEID3, GESLACHT, NATIVE, LEEFTIJD, OPLEID2, OPLEID1, LEEFGESL

b. Dependent Variable: RESPONS

Coefficients^a

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	28.981	1.584		18.292	.000
	LEEFTIJD	.215	.036	.299	6.024	.000
	GESLACHT	-3.090	1.828	-.269	-1.691	.094
	NATIVE	-6.020	.563	-.385	-10.692	.000
	LEEFGESL	-.170	.053	-.545	-3.239	.002
	OPLEID1	-12.277	.886	-1.048	-13.860	.000
	OPLEID2	-11.768	.888	-.978	-13.259	.000
	OPLEID3	-9.291	.958	-.635	-9.697	.000

a. Dependent Variable: RESPONS

- f. (5 punten) Hoeveel procent van de variantie van RESPONS kan met dit laatste model verklaard worden
- g. (5 punten) Hoe kun je nagaan of het de moeite waard is om de variabele OPLEID om te coderen in dummyvariabelen voordat deze in het model wordt opgenomen?

Vraag 2

Om preciezer inzicht te krijgen in de relatie tussen responstijd, vraaglengte (LENGTE) en vraagmoeilijkheid (MOEILIJK) voert de onderzoekster een variantieanalyse uit op alle 400 gegevens. Ga er vanuit dat de gegevens onafhankelijk zijn.

- a. (5 punten) Per respondent zijn 4 gegevens verzameld en het valt te verwachten dat dit een afhankelijkheid zal introduceren (al gaan we daar in dit tentamen dus NIET vanuit). Wat zou dan een betere analyse zijn?

Een variantieanalyse met als (fixed) factoren de lengte van de vraag (LENGTE: '0' is kort en '1' is lang) en de moeilijkheid van de vraag (MOEILIJK: '0' is makkelijk en '1' is moeilijk) en als afhankelijke variabele de responstijd (RESPONS) leverde het volgende op:

Univariate Analysis of Variance

Between-Subjects Factors

	Value	Label	N
LENGTE	0	kort	200
	1	lang	200
MOEILIJK	0	makkelijk	200
	1	moeilijk	200

Descriptive Statistics

Dependent Variable: RESPONS

LENGTE	MOEILIJK	Mean	Std. Deviation	N
0 kort	0 makkelijk	5.8892	4.84414	100
	1 moeilijk	10.7431	5.30054	100
	Total	8.3161	5.61879	200
1 lang	0 makkelijk	15.6900	5.76720	100
	1 moeilijk	20.7266	5.14724	100
	Total	18.2083	6.00841	200
Total	0 makkelijk	10.7896	7.23569	200
	1 moeilijk	15.7349	7.22499	200
	Total	13.2622	7.63387	400

Tests of Between-Subjects Effects

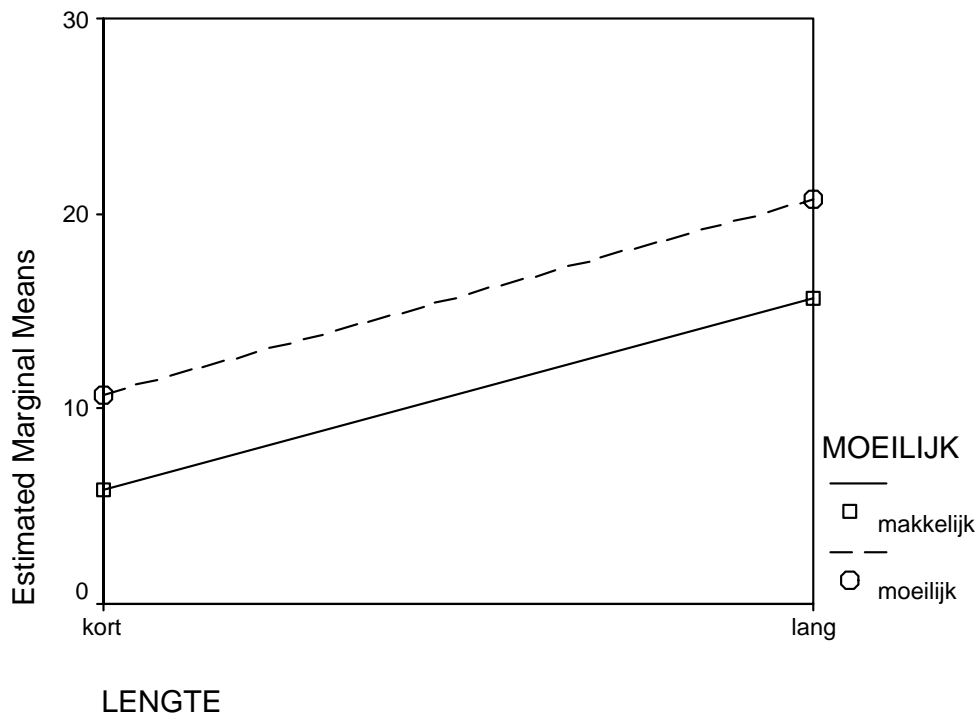
Dependent Variable: RESPONS

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	12231.847 ^a	3	4077.282	146.512	.000
Intercept	70354.645	1	70354.645	2528.103	.000
LENGTE	9785.463	1	9785.463	351.628	.000
MOEILIJK	2445.550	1	2445.550	87.878	.000
LENGTE * MOEILIJK	.834	1	.834	.030	.863
Error	11020.295	396	27.829		
Total	93606.787	400			
Corrected Total	23252.142	399			

a. R Squared = .526 (Adjusted R Squared = .522)

Profile Plots

Estimated Marginal Means of RESPONS



- b. (15 punten) Is er sprake van een significante interactie tussen LENGTE en MOEILIJK? Is er sprake van significante hoofdeffecten van LENGTE en/of MOEILIJK? En wat houdt dit in? Motiveer je antwoord!
- c. (5 punten) De bedoeling is dat we de responstijd kunnen gaan gebruiken om een idee van de nadentijd te kunnen krijgen. Dat kan eigenlijk alleen maar goed wanneer het lezen van de vraag en het nadenken over het antwoord twee processen zijn die na elkaar en onafhankelijk van elkaar in het hoofd van iedere respondent

plaatsvinden. Vormen de uitkomsten van dit onderzoek een bevestiging van dit idee of zijn ze in tegenspraak ermee? Motiveer je antwoord! Hint: denk eens na over de interactie...

Vraag 3

We zijn er vanuit gegaan dat er geen fouten gemaakt worden (verkeerde antwoorden gegeven worden). Een fout zou kunnen indiceren dat de respondent de vraag niet goed gelezen heeft, mogelijk omdat hij of zij te snel wilde lezen. In het experiment kreeg een respondent een nieuw setje van vier vragen indien hij of zij een fout maakte op één van de vragen uit het eerste setje. Alleen de responstijden op een set van vier correct beantwoorde vragen werden geregistreerd. Alles bij elkaar hebben 30% van de respondenten (minstens) één fout. Eerst gaan we analyseren welke factoren er aan bijdragen dat een persoon (minstens) een fout maakt in het eerst aangeboden setje vragen. We hebben dus 100 observaties!

Een onderzoeker begint met een kruistabel om te kijken of er een relatie is tussen het maken van een fout (FOUT: 0 is geen fout en 1 is minstens één fout) en of het Nederlands de eerste taal is (NATIVE). Dit kwam uit de analyse:

Crosstabs

NATIVE * FOUT Crosstabulation

				FOUT		Total
				0	1	
NATIVE	0 Nederlands niet eerste taal	Count	12	4	16	
		% within NATIVE	75.0%	25.0%	100.0%	
	1 Nederlands eerste taal	Count	58	26	84	
		% within NATIVE	69.0%	31.0%	100.0%	
Total		Count	70	30	100	
		% within NATIVE	70.0%	30.0%	100.0%	

- a. (5 punten) Wat is de oddsratio? Ongeacht of deze significant is: leidt het hebben van een andere taal dan het Nederlands als eerste taal tot meer fouten?

De onderzoeker besluit een stapsgewijze analyse te doen waarbij hij begint met de predictoren LEEFTIJD, GESLACHT, OPLEID, NATIVE. Kijk nog even bij opgave 1 wanneer je de coderingen niet meer helemaal zeker weet. Hij kiest voor een backward selectie op basis van de Likelihood Ratio test en als p-waarde voor de verwijdering van een term kiest hij voor 0.05. We gaan hier even voorbij aan de problemen die er mogelijk aan zo'n stapsgewijze analyse kleven (als een eerste exploratieve stap kan

het niet veel kwaad en 100 observaties is sowieso wat weinig voor een logistische regressieanalyse) en proberen de output te interpreteren.

Logistic Regression

Case Processing Summary

Unweighted Cases ^a		N	Percent
Selected Cases	Included in Analysis	100	100.0
	Missing Cases	0	.0
	Total	100	100.0
Unselected Cases		0	.0
Total		100	100.0

a. If weight is in effect, see classification table for the total number of cases.

Categorical Variables Codings

	Frequency	Parameter coding		
		(1)	(2)	(3)
OPLEID 1	40	1.000	.000	.000
2	35	.000	1.000	.000
3	19	.000	.000	1.000
4	6	.000	.000	.000

Block 1: Method = Backward Stepwise (Likelihood Ratio)

Omnibus Tests of Model Coefficients

	Chi-square	df	Sig.	
Step 1	Step	39.302	6	.000
	Block	39.302	6	.000
	Model	39.302	6	.000
Step 2 ^a	Step	-7.018	3	.071
	Block	32.284	3	.000
	Model	32.284	5	.000
Step 3 ^a	Step	-2.196	1	.138
	Block	30.087	2	.000
	Model	30.087	2	.000

a. A negative Chi-squares value indicates that the Chi-squares value has decreased from the previous step.

Variables in the Equation

	B	S.E.	Wald	df	Sig.	Exp(B)	95.0% C.I. for EXP(B)		
							Lower	Upper	
Step 1	LEEFTIJD	-.184	.052	12.741	1	.000	.832	.751	.920
	GESLACHT	3.024	.735	16.933	1	.000	20.583	4.874	86.925
	OPLEID			5.932	3	.115			
	OPLEID(1)	.571	1.156	.244	1	.621	1.770	.184	17.076
	OPLEID(2)	1.458	1.196	1.486	1	.223	4.296	.412	44.761
	OPLEID(3)	-.686	1.322	.269	1	.604	.504	.038	6.717
	NATIVE	1.446	.825	3.074	1	.080	4.246	.843	21.378
	Constant	1.369	1.785	.588	1	.443	3.930		
Step 2	LEEFTIJD	-.166	.046	12.803	1	.000	.847	.773	.928
	GESLACHT	2.705	.651	17.244	1	.000	14.954	4.172	53.610
	NATIVE	1.077	.757	2.025	1	.155	2.935	.666	12.933
	Constant	1.978	1.351	2.144	1	.143	7.229		
Step 3	LEEFTIJD	-.152	.045	11.531	1	.001	.859	.787	.938
	GESLACHT	2.598	.630	16.978	1	.000	13.432	3.904	46.213
	Constant	2.488	1.297	3.679	1	.055	12.037		

a. Variable(s) entered on step 1: LEEFTIJD, GESLACHT, OPLEID, NATIVE.

Model if Term Removed

Variable	Model Log Likelihood	Change in -2 Log Likelihood	df	Sig. of the Change	
Step 1	LEEFTIJD	-50.559	18.247	1	.000
	GESLACHT	-54.085	25.298	1	.000
	OPLEID	-44.944	7.018	3	.071
	NATIVE	-43.119	3.367	1	.067
Step 2	LEEFTIJD	-53.873	17.857	1	.000
	GESLACHT	-56.791	23.692	1	.000
	NATIVE	-46.043	2.196	1	.138
Step 3	LEEFTIJD	-53.975	15.864	1	.000
	GESLACHT	-57.421	22.756	1	.000

Variables not in the Equation

	Score	df	Sig.		
Step 2	Variables	OPLEID	6.546	3	.088
		OPLEID(1)	.063	1	.801
		OPLEID(2)	4.786	1	.029
		OPLEID(3)	3.905	1	.048
	Overall Statistics		6.546	3	.088
Step 3	Variables	OPLEID	5.553	3	.135
		OPLEID(1)	.201	1	.654
		OPLEID(2)	4.418	1	.036
		OPLEID(3)	3.068	1	.080
		NATIVE	2.093	1	.148
	Overall Statistics		8.362	4	.079

a. Variable(s) removed on step 2: OPLEID.

b. Variable(s) removed on step 3: NATIVE.

- b. (10 punten) Geeft de formule van het best passende regressiemodel. Geef ook een interpretatie aan de coëfficiënten (voor eventuele continue variabelen: gaat het aantal fouten omhoog of omlaag als de variabele groter wordt; voor discrete variabelen: welke uitkomst heeft het meeste fouten).

Mocht een respondent die een fout maakt in het eerste setje vragen en dus een tweede setje krijgt, ook daar weer een fout in maken dan volgt een derde setje etc. Zo kun je dus ook kijken naar het aantal extra setjes dat een respondent heeft moeten beantwoorden voordat een foutvrij setje antwoorden werd geproduceerd. Op deze nieuwe variabele scoren dus 70 personen een '0' (want zij maakten geen fouten) etc. Het blijkt dat deze nieuwe variabele goed met behulp van Poissonregressie gemodelleerd kan worden. De gekozen linkfunctie is de log-functie. De enige significante predictor is GESLACHT (0 voor mannen en 1 voor vrouwen). De waarde van het regressiegewicht voor geslacht is 0.70.

- c. (5 punten) Hoe verhouden zich het aantal extra setjes dat mannen nodig hebben en dat vrouwen nodig hebben tot elkaar?

Vraag 4

- a. (7 punten) Bij een seizoensdecompositie van een tijdreeks (2001 t.m. 2003) van de hoeveelheid UV-straling (in een willekeurige meeteenheid gemeten op zonnige dagen op een vaste testsite in De Bilt op het moment dat de zon zo hoog mogelijk staat) per is er per ongeluk een blikje cola op de uitvoer gekomen. Vandaar dat enige getallen zijn weggefallen. Vul de lege plekken in indien de data dat toestaan en geef aan hoe je aan je getallen komt. De decompositie is met behulp van een multiplicatief model uitgevoerd. Vul je getallen in op het inlevervel en zet ofwel op het inlevervel ofwel op je uitwerkingen hoe je aan de getallen komt!!

NB de notatie van de hand-out is aangehouden!

t	tijd	L	Y_t	U_t	$S_t * O_t$	S_t	G_t	T_t	$C_t * O_t$	C_t	O_t
1	2001	1	19			1.079	17.609		1.004		
2	2001	2	61					22.615	1.101	1.006	1.094
3	2001	3	8		0.300	0.316	25.316	27.691	0.914	1.065	0.858
4	2001	4	6	34.000	0.176	0.155	38.710	32.767	1.181	1.058	1.116
5	2002	1	44	39.000	1.128	1.079	40.778	37.843	1.078		
6	2002	2	95	39.875	2.382			42.919	0.903	0.968	0.933
7	2002	3	14	42.375	0.330	0.316	44.304	47.995	0.923	0.892	1.035
8	2002	4	7	52.625		0.155	45.161	53.071		0.926	0.919
9	2003	1	63	61.500	1.024	1.079	58.387	58.147	1.004	0.958	1.048
10	2003	2	158	63.125	2.503			63.223	1.020	1.014	1.006
11	2003	3	22			0.316	69.620	68.299	1.019	1.031	0.988
12	2003	4	12			0.155	77.419	73.375	1.055		

Hierbij is nog de volgende uitkomst gegeven:

De op basis van G_t geschatte trendlijn is: $Y_t = 12.463 + 5.076 \times t$

- b. (3 punten) Dit jaar is er een extreem hoge hoeveelheid UV gemeten. Dit heeft waarschijnlijk te maken gehad met de extreme koude boven de Noordpool van afgelopen winter. In welke term (seizoensinvloed, trend, conjunctuur, onregelmatig) zou deze extreme waarde zich openbaren? Motiveer je antwoord.

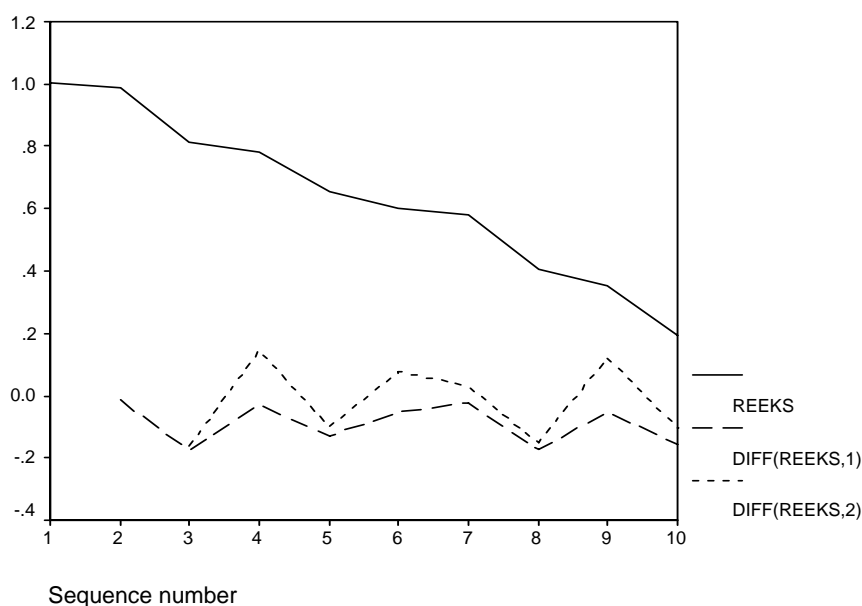
Vraag 5

Iedere statisticus zal wijzen op het gevaar om gesofisticeerde tijdreeksmodellen (zoals b.v. ARIMA modellen) op te korte tijdreeksen toe te passen. Dat kan namelijk al snel leiden tot verkeerde modelidentificaties. Een potentiële bron van verwarring is aannemen dat de tijdreeks niet stationair is (en er dus gedifferentieerd moet worden) terwijl er eigenlijk sprake is van een autoregressief (AR) model. Immers bij een AR model is er sprake van een samenhang met de observatie van het tijdstip ervoor.

- a. (3 punten) Stel dat we te maken hebben met het volgende AR model:

$Y_t = 0.95 \times Y_{t-1} + Z_t$, en dat Z_t een hele kleine variantie heeft. Om een idee te krijgen van wat er gebeurt, zetten we deze variantie even op nul. Schets (voor t is 1 tm 10) hoe Y_t als functie van t verloopt wanneer Y_0 gelijk is aan 1.

In het volgende voorbeeld is REEKS gegenereerd volgens een AR(1) model (dus zonder integreren, het omgekeerde van verschillen) en zou dus stationair moeten zijn.



Descriptive Statistics

	N	Minimum	Maximum	Mean	Std. Deviation
REEKS	10	.20	1.00	.6368	.26718
REEKS_1 DIFF(REEKS,1)	9	-.18	-.01	-.0896	.06802
REEKS_2 DIFF(REEKS,2)	8	-.16	.15	-.0179	.12519
Valid N (listwise)	8				

- b. (4 punten) Als je kijkt naar bovenstaande output, welke reeks zou je dan stationair noemen: de oorspronkelijke (REEKS), de éénmaal gedifferentieerde (REEKS_1) of de tweemaal gedifferentieerde reeks (REEKS_2). Motiveer je antwoord!

Hieronder staan stukken van de output van een ARIMA(1,0,0) (dus een AR1 model), een ARIMA(0,1,0) (dus alleen differencing) en een ARIMA(1,1,0) model.

ARIMA(1,0,0):

Model Description:

Variable: REEKS

Regressors: NONE

Non-seasonal differencing: 0

No seasonal component in model.

Parameters:

AR1 _____ < value originating from estimation >

CONSTANT _____ < value originating from estimation >

FINAL PARAMETERS:

Number of residuals 10
 Standard error .12252994
 Log likelihood 6.727502
 AIC -9.455004
 SBC -8.8498338

Analysis of Variance:

	DF	Adj. Sum of Squares	Residual Variance
Residuals	8	.15245131	.01501359

Variables in the Model:

	B	SEB	T-RATIO	APPROX. PROB.
AR1	.95281861	.08138153	11.708044	.00000259
CONSTANT	.60604751	.36584832	1.656554	.13620055

ARIMA(0,1,0)

Model Description:

Variable: REEKS

Regressors: NONE

Non-seasonal differencing: 1

No seasonal component in model.

Parameters:

CONSTANT _____ < value originating from estimation >

FINAL PARAMETERS:

Number of residuals	9
Standard error	.06801922
Log likelihood	11.921238
AIC	-21.842476
SBC	-21.645251

Analysis of Variance:

	DF	Adj. Sum of Squares	Residual Variance
Residuals	8	.03701292	.00462661

Variables in the Model:

	B	SEB	T-RATIO	APPROX. PROB.
CONSTANT	-.08955946	.02267307	-3.9500358	.00423581

ARIMA(1,1,0)

Model Description:

Variable: REEKS

Regressors: NONE

Non-seasonal differencing: 1

No seasonal component in model.

Parameters:

AR1 _____ < value originating from estimation >

CONSTANT _____ < value originating from estimation >

FINAL PARAMETERS:

Number of residuals	9
Standard error	.04696593
Log likelihood	15.370802
AIC	-26.741603
SBC	-26.347154

Analysis of Variance:

	DF	Adj. Sum of Squares	Residual Variance
Residuals	7	.01713353	.00220580

Variables in the Model:

	B	SEB	T-RATIO	APPROX. PROB.
AR1	-.77970462	.22099196	-3.5282036	.00962101
CONSTANT	-.08997809	.00925882	-9.7180904	.00002582

- c. (3 punten) Welke van bovenstaande modellen zou je op basis van deze analyses als best passend model aanwijzen?