Vraag 1

Bij een enquête die met behulp van 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 is te definiëren als de tijd die een respondent er over doet om zijn antwoord aan te geven. Deze responstijd kan geanalyseerd worden om zo extra informatie te verzamelen. Zo kan het zijn dat vragen waarop snel geantwoord wordt, indiceren dat dit vragen over kwesties zijn waarbij de respondent overtuigd is van zijn of haar mening, terwijl vragen waarover langer wordt nagedacht gaan over kwesties waar de respondent minder zeker van is. Zo kan het voor de waarschijnlijkheid dat iemand echt overweegt om het openbaar vervoer als een alternatief voor de auto te beschouwen, nogal wat uitmaken of deze persoon lang of kort nadenkt over de vraag "Het openbaar vervoer is een bruikbaar alternatief voor de auto", ook al geven ze allebei hetzelfde antwoord!
In de responstijd zit echter ook de tijd die het kost om een vraag te lezen en te begrijpen. Er kunnen allerlei factoren van invloed zijn op deze tijd die niets met de mening van de respondent te maken heeft. Denk bijvoorbeeld aan leeftijd, opleiding, geslacht, etniciteit etc. Om een beeld te krijgen van de mogelijke invloeden van deze factoren wordt een experiment uitgevoerd, waaraan 200 vrijwilligers meedoen.
Op een computerscherm wordt een plaatje geprojecteerd met onder dat plaatje een vraag die over het plaatje gaat.
Een onderzoekster

Bij de beoordeling van de kwaliteit van een nieuw ontwikkeld type kunstgras worden gedurende een experimentele periode van 1 jaar in verschillende praktijktesten gegevens verzameld. De afhankelijke variabele is slijtage (SLIJTAGE), uitgedrukt in een percentage; 0% is geen slijtage en 100% houdt in dat het kunstgras niet meer te gebruiken is. De slijtage wordt per vierkante meter bepaald en dan gemiddeld over het hele oppervlakte. Als voorspellende variabelen worden meegenomen: type sport dat erop beoefend wordt (SPORT), aantal zonuren waaraan het kunstgras is blootgesteld geweest in de experimentele periode (ZON), minimale temperatuur waaraan het kunstgras is blootgesteld geweest (MINTEMP), maximale temperatuur waaraan het kunstgras is blootgesteld geweest (MAXTEMP), de hoeveelheid neerslag waaraan het kunstgras is blootgesteld geweest (NEERSLAG) en de bezettingsgraad van het terrein, d.w.z. het percentage van de tijd dat het terrein gebruikt is (BEZET).

Hieronder staat de output van een regressieanalyse

Descriptive Statistics

	Mean	Std. Deviation	N
SLIJTAGE	27.9218	13.39049	310
SPORT	3.60	1.739	310
ZON	1001.5899	198.79043	310
MINTEMP	-6.9492	2.00521	310
MAXTEMP	25.0454	2.08384	310
NEERSLAG	740.5536	4.87789	310
BEZET	42.9680	13.10845	310

Variables Entered/Removed[b]

Model	Variables Entered	Variables Removed	Method
1	BEZET, NEERSLAG, ZON, MINTEMP, MAXTEMP, SPORT[a]	.	Enter

a. All requested variables entered.

b. Dependent Variable: SLIJTAGE

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.604[a]	.365	.353	10.77487

a. Predictors: (Constant), BEZET, NEERSLAG, ZON, MINTEMP, MAXTEMP, SPORT

ANOVA[b]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	20227.697	6	3371.283	29.038	.000[a]
	Residual	35177.633	303	116.098		
	Total	55405.330	309			

a. Predictors: (Constant), BEZET, NEERSLAG, ZON, MINTEMP, MAXTEMP, SPORT

b. Dependent Variable: SLIJTAGE

Coefficients[a]

Model		Unstandardized Coefficients B	Std. Error	Standardized Coefficients Beta	t	Sig.	95% Confidence Interval for B Lower Bound	Upper Bound
1	(Constant)	114.255	93.749		1.219	.224	-70.226	298.736
	SPORT	.156	.395	.020	.395	.693	-.621	.934
	ZON	.014	.003	.208	4.487	.000	.008	.020
	MINTEMP	.344	.306	.051	1.122	.263	-.259	.947
	MAXTEMP	-.283	.296	-.044	-.956	.340	-.865	.299
	NEERSLAG	-.157	.126	-.057	-1.244	.214	-.404	.091
	BEZET	.570	.052	.558	10.952	.000	.468	.673

a. Dependent Variable: SLIJTAGE

a. (5 punten) Stel dat de bezettingsgraad van een terrein 10% stijgt (b.v. van 20% naar 30% bezet) terwijl alle andere factoren gelijk blijven. Wat is dan de toename of afname in de slijtage. Motiveer je antwoord.

b. (5 punten) Het intercept heeft de waarde 114.255 terwijl de afhankelijke variabele een percentage is en niet hoger dan 100 kan worden. Is de schatting van het intercept wel te vertrouwen?

c. (5 punten) Kun je op basis van deze analyse stellen dat er sprake is van een significant effect van het type sport dat op het veld beoefend wordt. Motiveer je antwoord.

d. (5 punten) Hoe zou je met behulp van een regressiemodel de hypothese kunnen onderzoeken dat het effect van de bezettingsgraad voor verschillende sporten anders uitpakt (m.a.w. dat het effect van meer korfbal op een terrein verschilt van het effect van meer voetbal op het terrein)?

e. (5 punten) Stel dat er op basis van ander onderzoek bekend is dat de hoeveelheid neerslag een erg belangrijke factor is op de slijtage. Hoe valt deze conclusie met dit onderzoek te rijmen? Anders gezegd: wat zou een (statistische) verklaring kunnen zijn voor het feit dat in dit onderzoek geen effect van de hoeveelheid neerslag gevonden wordt, maar wel van andere factoren zoals de hoeveelheid zon en de bezettingsgraad?

Vraag 2 (10 punten)

Bij een seizoensdecompositie van een tijdreeks (2000 t.m. 2002) van het aantal hectare kunstgras ten behoeve van sportterreinen per kwartaal is er wederom per ongeluk een kop koffie op de uitvoer gekomen. Vandaar dat enige getallen zijn weggevallen. 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.

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	2000	1	16			1.307	12.245	2.773	4.416		
2	2000	2	14			1.046	13.390	7.594	1.763	2.457	0.718
3	2000	3	13	15.625	0.832			12.415	1.193	1.337	0.892
4	2000	4	14	18.625	0.752	0.770	18.177	17.236	1.055	1.062	0.993
5	2001	1	27	21.250	1.271	1.307	20.663	22.057	0.937	0.984	0.952
6	2001	2	27	22.875	1.180	1.046	25.823	26.878	0.961	0.884	1.087
7	2001	3	21	24.500	0.857			31.699	0.755		
8	2001	4	19	26.000		0.770	24.669	36.520	0.676	0.693	0.975
9	2002	1	35	28.125	1.244	1.307	26.786	41.341	0.648	0.655	0.989
10	2002	2	31		0.832	1.046	29.648	46.162	0.642	0.683	0.940
11	2002	3	34					50.983		1.080	0.704
12	2002	4	79			0.770	102.573		1.838		

Hierbij is nog de volgende uitkomst gegeven:

De geschatte trendlijn is: $Y_t = -2.0479 + 4.821 \times t$

Vraag 3

Ook in een laboratoriumexperiment wordt de slijtvastheid van het nieuwe kunstgras onderzocht. Op een kunstmatige wijze wordt gedurende een uur het kunstgras tot het uiterste belast. Dan wordt naar de slijtage gekeken. De slijtage wordt uitgedrukt als promillage achteruitgang in kwaliteit ten opzichte van de vorige meting (NB je had natuurlijk ook telkens de kwaliteit kunnen bepalen maar men heeft hiervoor gekozen). Vervolgens wordt hetzelfde stuk kunstgras weer een uur belast en wordt weer de slijtage gemeten. Zo worden 200 metingen aan dit éne stuk kunstgras verricht.

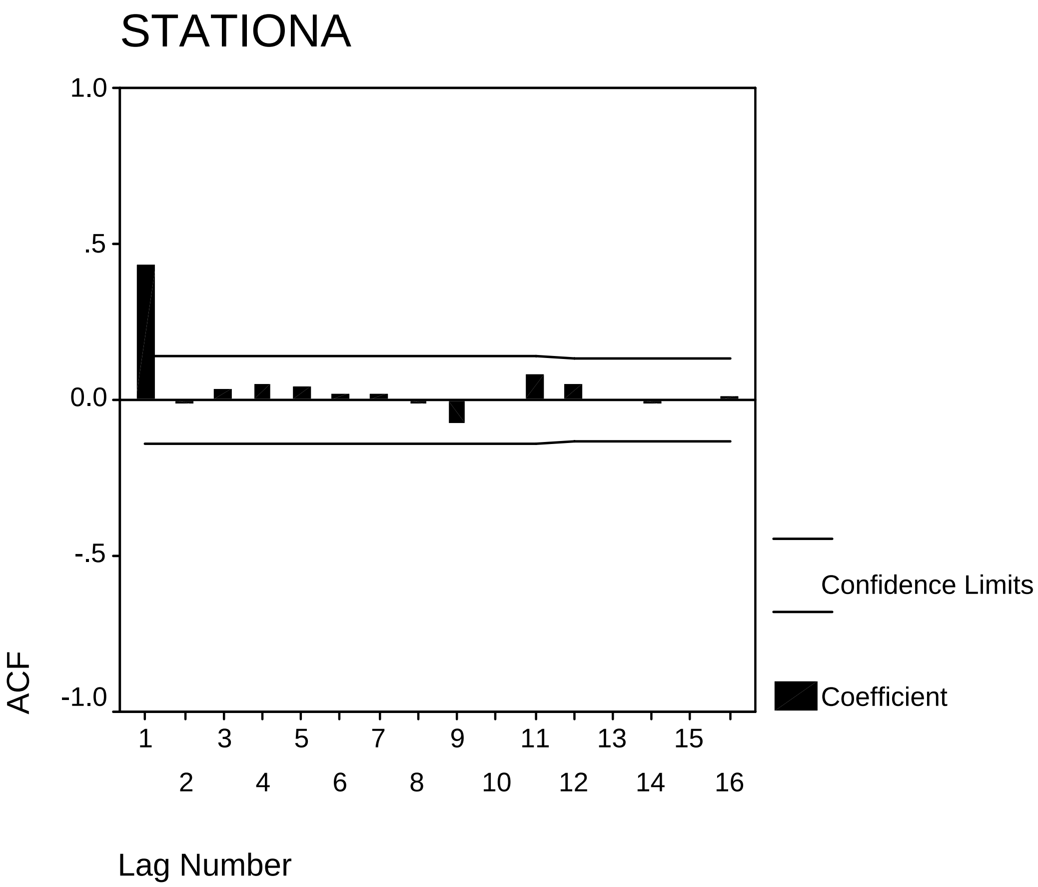
In onderstaande tabel staan wat beschrijvende maten van deze gegevens (SLIJTAGE) en van de gegevens na éénmaal differencing (met lag 1; DIF(SLIJTAGE,1)) en na tweemaal differencing (DIF(SLIJTA_1,1)).

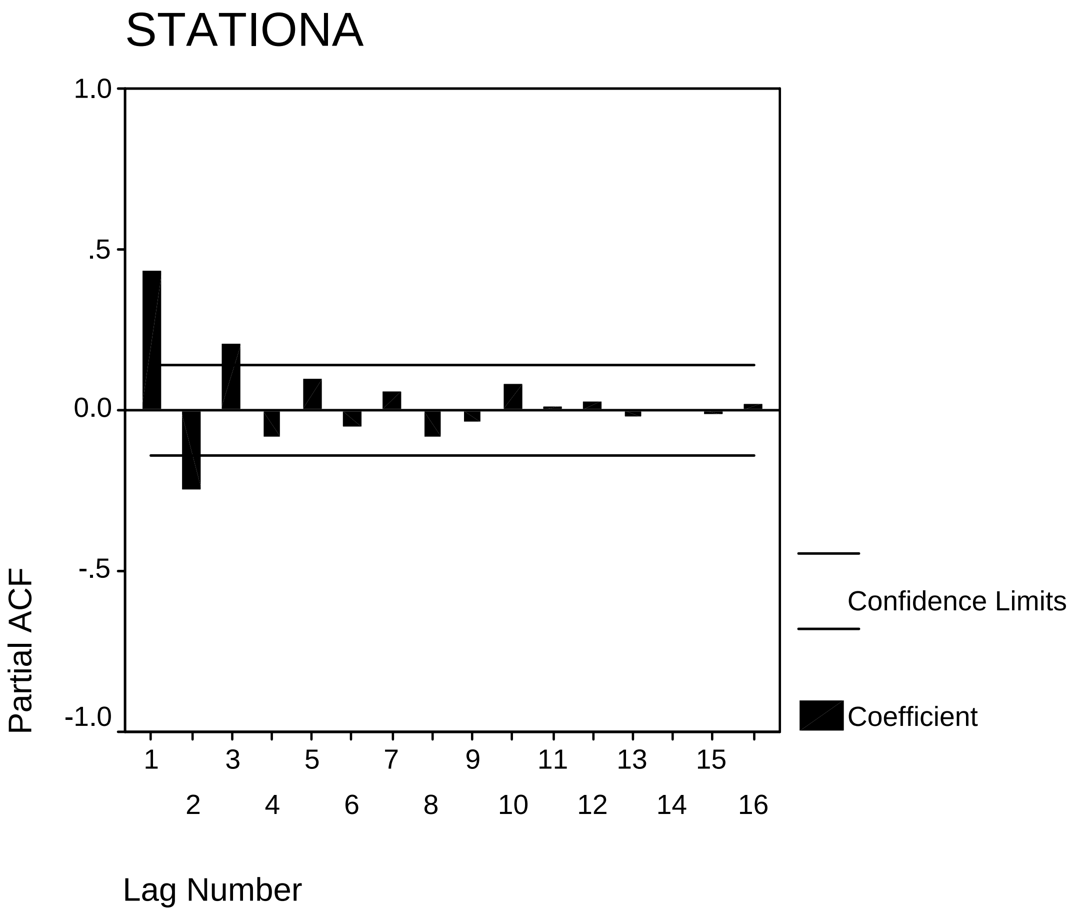
Descriptive Statistics

	N	Minimum	Maximum	Mean	Std. Deviation	Variance
SLIJTAGE	200	1.92	7.44	4.8102	1.23160	1.517
DIFF(SLIJTAGE,1)	199	-3.48	3.54	-.0132	1.30444	1.702
DIFF(SLIJTA_1,1)	198	-5.66	5.20	.0049	1.93712	3.752

a. (3 punten)Wat is de stationaire reeks en waar zie je dat aan?

Deze stationaire reeks krijgt een nieuwe naam (STATIONA) wordt verder geanalyseerd. Eerst wordt gekeken naar de autocorrelaties en partiële autocorrelaties:





b. (2 punten) Welke orde autocorrelatie is nog significant? Waaraan zie je dat?

Op zich kun je aan deze plaatjes al bepalen wat voor type ARIMA model er op de stationaire reeks te passen is, maar omdat de onderzoekers niet zeker van hun zaak zijn passen ze zowel een AR(1) als een MA(1) als een ARMA(1,1) model op de stationaire reeks.

Hieronder staan delen van de output van de drie modellen:

Van het AR(1) model:

```
Model Description:
Variable:   STATIONA
Regressors: NONE

Non-seasonal differencing: 0
No seasonal component in model.

Parameters:
AR1       ________  < value originating from estimation >
CONSTANT ________  < value originating from estimation >
95.00 percent confidence intervals will be generated.
```

```
FINAL PARAMETERS:


Standard error       1.1099755
Log likelihood       -303.76122
AIC                  611.52244
SBC                  618.11907

            Analysis of Variance:

               DF  Adj. Sum of Squares    Residual Variance

Residuals                  244.20605            1.2320455

           Variables in the Model:

                    B          SEB     T-RATIO   APPROX. PROB.

AR1           .4388314   .06380830    6.877340        .0000000
CONSTANT     4.8149452   .13932002   34.560325        .0000000
```

Van het MA(1) model:

```
Model Description:
Variable:   STATIONA
Regressors: NONE

Non-seasonal differencing: 0
No seasonal component in model.

Parameters:
MA1      ________  < value originating from estimation >
CONSTANT ________  < value originating from estimation >
95.00 percent confidence intervals will be generated.

FINAL PARAMETERS:


Standard error       1.0411953
Log likelihood       -291.14246
AIC                  586.28492
SBC                  592.88156

            Analysis of Variance:

               DF  Adj. Sum of Squares    Residual Variance

Residuals                  215.25836            1.0840877

           Variables in the Model:

                     B          SEB      T-RATIO   APPROX. PROB.

MA1           -.6576990   .05398848   -12.182209        .0000000
CONSTANT      4.8138748   .12180442    39.521347        .0000000
```

En tenslotte van het ARMA(1,1) model:

```
Model Description:
Variable:   STATIONA
Regressors: NONE

Non-seasonal differencing: 0
No seasonal component in model.

Parameters:
AR1      ________  < value originating from estimation >
MA1      ________  < value originating from estimation >
CONSTANT ________  < value originating from estimation >
95.00 percent confidence intervals will be generated.

FINAL PARAMETERS:

Standard error       1.0432948
Log likelihood       -291.04315
AIC                  588.0863
SBC                  597.98125

            Analysis of Variance:

               DF  Adj. Sum of Squares    Residual Variance

Residuals                    215.03411            1.0884641

           Variables in the Model:

                     B         SEB      T-RATIO   APPROX. PROB.

AR1          -.0472902   .10863033     -.435332       .66379824
MA1          -.6830901   .08017559    -8.519926       .00000000
CONSTANT     4.8135409   .11834516    40.673744       .00000000
```

c. (10 punten) Wat is het bestpassende model? Geef aan hoe je tot die conclusie komt. Geef ook de formule die bij het model hoort

Vraag 4

Naast de kwaliteit van het kunstgras zelf, wordt er ook nog onderzoek gedaan naar nieuwe manieren om het kunstgras te leggen. Vooral de ondergrond is hierbij van belang. In een onderzoek worden drie verschillende manieren van het prepareren van de ondergrond met elkaar vergeleken; voor het gemak noemen we deze methodes A, B en C. Er worden 12 sportterreinen geselecteerd en deze worden op een willekeurige manier in drie groepen van vier terreinen elk verdeeld. In iedere groep wordt de ondergrond met behulp van één van de methodes geprepareerd. De terreinen worden vervolgens een jaar lang gebruikt en dan wordt wederom de slijtage gemeten, waarbij dezelfde meetprocedure als in vraag 1 gebruikt wordt.

De data staan in onderstaande tabel, gevolgd door uitkomsten van een op deze data uitgevoerde analyse.

A	B	C
26.86	29.97	32.13
33.15	34.40	35.68
19.35	20.03	22.83
16.78	20.41	22.07

Oneway

Descriptives

SLIJTAGE

	N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean		Minimum	Maximum
					Lower Bound	Upper Bound		
A	4	24.0352	7.42955	3.71478	12.2131	35.8573	16.78	33.15
B	4	26.1998	7.14374	3.57187	14.8325	37.5671	20.03	34.40
C	4	28.1754	6.77990	3.38995	17.3871	38.9638	22.07	35.68
Total	12	26.1368	6.68038	1.92846	21.8923	30.3813	16.78	35.68

ANOVA

SLIJTAGE

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	34.307	2	17.154	.338	.722
Within Groups	456.595	9	50.733		
Total	490.902	11			

Post Hoc Tests

Multiple Comparisons

Dependent Variable: SLIJTAGE
Scheffe

(I) METHODE	(J) METHODE	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
A	B	-2.1646	5.03651	.913	-16.8596	12.5304
	C	-4.1403	5.03651	.722	-18.8353	10.5548
B	A	2.1646	5.03651	.913	-12.5304	16.8596
	C	-1.9757	5.03651	.927	-16.6707	12.7194
C	A	4.1403	5.03651	.722	-10.5548	18.8353
	B	1.9757	5.03651	.927	-12.7194	16.6707

a. (5 punten) Kan op basis van deze analyse gesteld worden dat er een verschil zit tussen de drie verschillende methodes? Waaraan lees je dat af?

Het bovengenoemde verhaal bleek bij nadere beschouwing niet helemaal correct. Er waren 3 voetbalvelden, 3 rugbyvelden, 3 korfbalvelden en 3 honkbalvelden en van ieder type sportveld is er één (op willekeurige wijze gekozen) aan iedere methode toegewezen. Dit feit is bij de eerste analyse niet meegenomen en wordt nu bij een vervolg analyse wel meegenomen. De tabel met de data en de uitkomsten van de analyse staan hieronder vermeld.

Sport	A	B	C
Voetbal	26.86	29.97	32.13
Rugby	33.15	34.40	35.68
Korfbal	19.35	20.03	22.83
Honkbal	16.78	20.41	22.07

Univariate Analysis of Variance

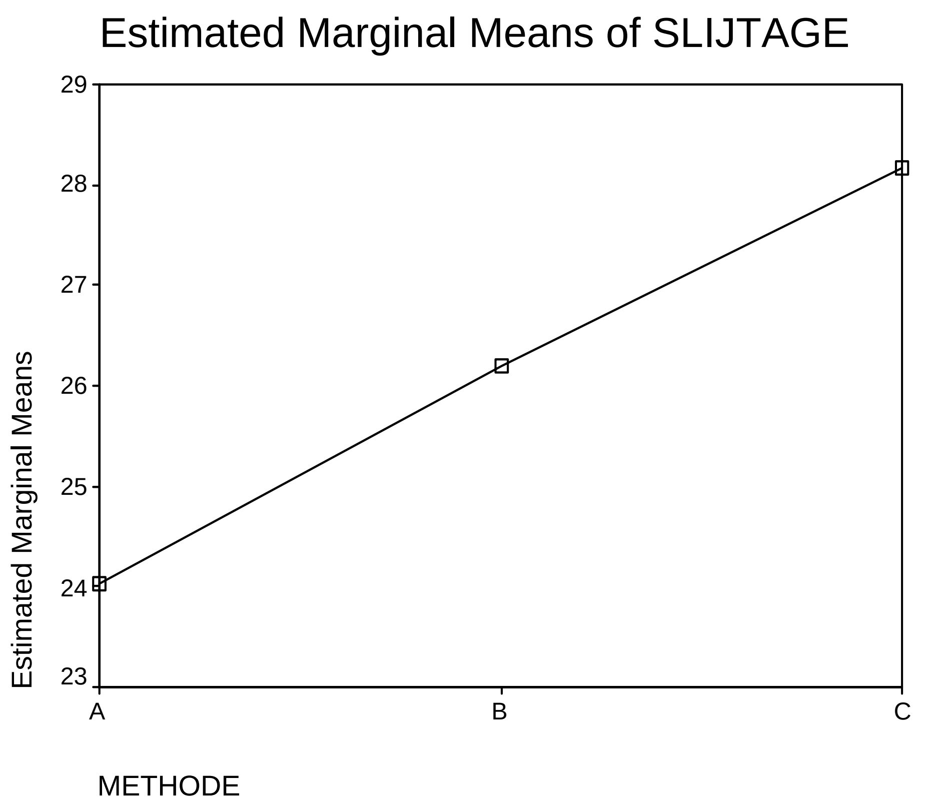
Tests of Between-Subjects Effects

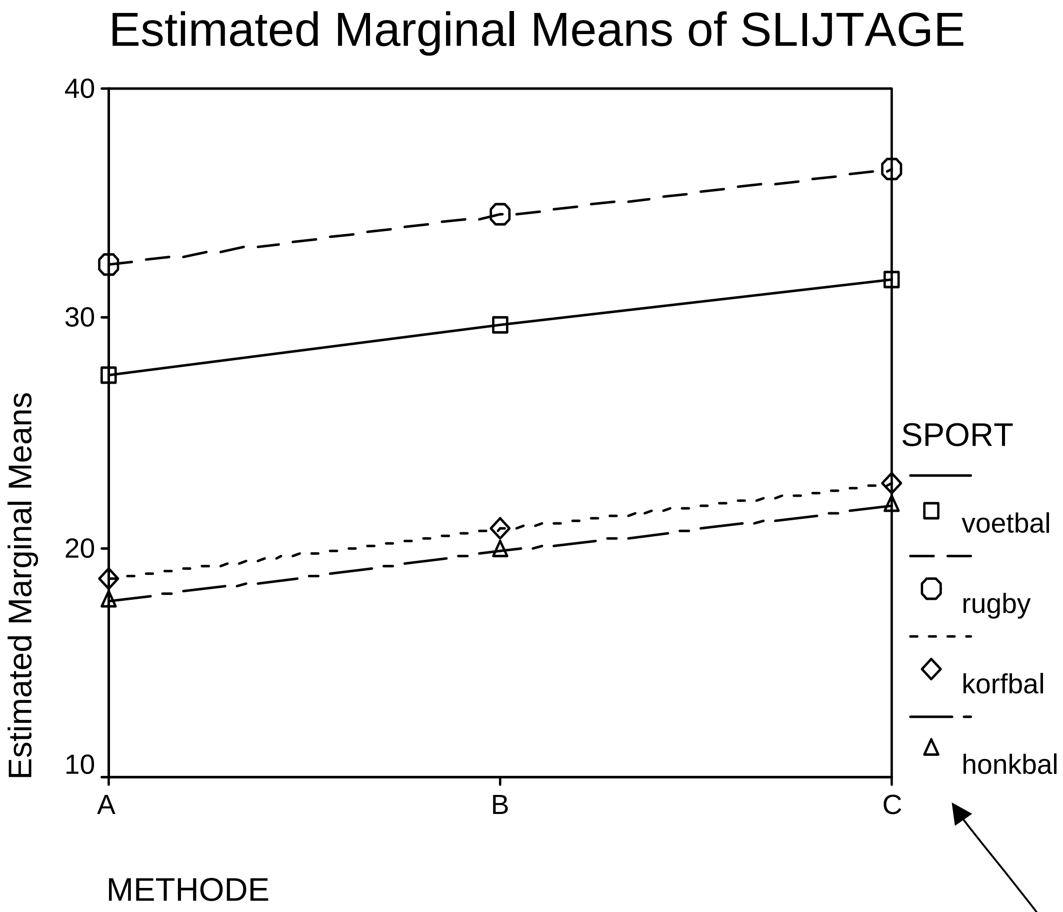
Dependent Variable: SLIJTAGE

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	486.570[a]	5	97.314	134.786	.000
Intercept	8197.589	1	8197.589	11354.153	.000
METHODE	34.307	2	17.154	23.759	.001
SPORT	452.263	3	150.754	208.804	.000
Error	4.332	6	.722		
Total	8688.491	12			
Corrected Total	490.902	11			

a. R Squared = .991 (Adjusted R Squared = .984)

Profile Plots





b. (5 punten) Kun je op basis van deze analyse stellen dat er een verschil is tussen de methodes? Waaraan lees je dat af?

c. (5 punten) Stel 95% betrouwbaarheidsintervallen op voor de verschillen tussen de methodegemiddeldes. Wanneer je deze betrouwbaarheidintervallen vergelijkt met de intervallen berekend in de vorige analyse (zie tabel **Post Hoc Tests**) wat valt je dan op?

d. (5 punten) In het tweede plaatje staan parallelle lijnen voor de vier sporten. Dit komt niet overeen met de geobserveerde getallen. Dit plaatje geeft namelijk de op basis van het gehanteerde model geschatte gemiddeldes. Leg uit waarom deze lijnen parallel lopen.

e. (5 punten) Stel je zou de hypothese willen onderzoeken dat bepaalde methodes voor preparatie van de ondergrond beter geschikt zijn voor de éne sportsoort dan voor de andere (m.a.w. dat je b.v. beter methode A kunt gebruiken voor voetbalvelden en methode C voor honkbalvelden). Hoe zou je je experiment dan moeten opzetten en welke analyse zou je moeten gebruiken?

Vraag 5

Er is veel te doen over het feit dat sporten op kunstgras tot meer blessures bij de sporters zou leiden. Om dit te onderzoeken zijn op een aantal willekeurig gekozen dagen bij zowel een willekeurig gekozen aantal kunstgras velden als bij een willekeurig gekozen aantal ‘echte’ grasvelden bijgehouden of er één of meer ernstige blessures zijn gevallen. De afhankelijke variabele die geanalyseerd wordt is BLESSURE waarbij een 0 staat voor geen blessures en een 1 voor wel een of meer blessures. De onafhankelijke variabele die we in eerste instantie bekijken is KUNSTGRA (0 = echt gras, 1 = kunstgras).
Hieronder staan de resultaten van een aantal eerste analyses.

KUNSTGRA * BLESSURE Crosstabulation

Count

		BLESSURE		
		.00	1.00	Total
KUNSTGRA	.00	122	28	150
	1.00	97	53	150
Total		219	81	300

Logistic Regression

Dependent Variable Encoding

Original Value	Internal Value
.00	0
1.00	1

Block 0: Beginning Block

Classification Table[a,b]

	Observed		Predicted		
			BLESSURE		Percentage Correct
			.00	1.00	
Step 0	BLESSURE	.00	219	0	100.0
		1.00	81	0	.0
	Overall Percentage				73.0

a. Constant is included in the model.

b. The cut value is .500

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 0	Constant	-.995	.130	58.496	1	.000	.370

Variables not in the Equation

			Score	df	Sig.
Step 0	Variables	KUNSTGRA	10.570	1	.001
	Overall Statistics		10.570	1	.001

Block 1: Method = Enter

Omnibus Tests of Model Coefficients

		Chi-square	df	Sig.
Step 1	Step	10.704	1	.001
	Block	10.704	1	.001
	Model	10.704	1	.001

Model Summary

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	339.252	.035	.051

Classification Table[a]

Observed			Predicted		
			BLESSURE		Percentage Correct
			.00	1.00	
Step 1	BLESSURE	.00	219	0	100.0
		1.00	81	0	.0
	Overall Percentage				73.0

a. The cut value is .500

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)	95.0% C.I.for EXP(B)	
								Lower	Upper
Step 1[a]	KUNSTGRA	.867	.270	10.294	1	.001	2.381	1.401	4.044
	Constant	-1.472	.210	49.333	1	.000	.230		

a. Variable(s) entered on step 1: KUNSTGRA.

a. (5 punten) Zou je op basis van deze analyse concluderen dat er een relatie is tussen het spelen op kunstgras en het oplopen van een blessure?

b. (5 punten) Hoe ziet die relatie eruit, met andere woorden: is het gevaarlijker om op kunstgras te spelen of is dat juist niet zo? Waaraan zie je dat?

Nu zijn er allerlei mogelijke verstorende variabelen. Op kunstgras kan vaker gespeeld worden, ook onder omstandigheden (b.v. vorst of neerslag) die het spelen op gewoon gras belemmeren. Wanneer het nu ook zo is dat er een relatie is tussen deze omstandigheden en het oplopen van een blessure dan kan het zijn dat er allerlei valse verbanden gevonden worden. Daarom worden de volgende correctiefactoren in het model opgenomen: of er al dan niet sprake was van neerslag tijdens de wedstrijd

(NEERSLAG), of het al dan niet vroor tijdens de wedstrijd (VORST), de interactie tussen deze twee variabelen, de gespeelde sport (SPORT) en het tijdstip waarop de wedstrijd begon (TIJD). Deze variabelen worden in blok 1 in het model opgenomen en in blok 2 volgt dan KUNSTGRA De uitkomsten van de analyse zijn als volgt.

Logistic Regression

Case Processing Summary

Unweighted Cases[a]		N	Percent
Selected Cases	Included in Analysis	300	100.0
	Missing Cases	0	.0
	Total	300	100.0
Unselected Cases		0	.0
Total		300	100.0

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

Dependent Variable Encoding

Original Value	Internal Value
.00	0
1.00	1

Categorical Variables Codings

		Frequency	Parameter coding		
			(1)	(2)	(3)
SPORT	voetbal	75	1.000	.000	.000
	rugby	75	.000	1.000	.000
	korfbal	75	.000	.000	1.000
	honkbal	75	.000	.000	.000

Block 0: Beginning Block

Classification Table[a,b]

	Observed		Predicted		
			BLESSURE		Percentage Correct
			.00	1.00	
Step 0	BLESSURE	.00	219	0	100.0
		1.00	81	0	.0
	Overall Percentage				73.0

a. Constant is included in the model.

b. The cut value is .500

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 0	Constant	-.995	.130	58.496	1	.000	.370

Variables not in the Equation

			Score	df	Sig.
Step 0	Variables	VORST	11.594	1	.001
		NEERSLAG	31.018	1	.000
		NEERSLAG by VORST	33.909	1	.000
		SPORT	4.651	3	.199
		SPORT(1)	2.982	1	.084
		SPORT(2)	2.486	1	.115
		SPORT(3)	.276	1	.599
		TIJD	1.887	1	.170
	Overall Statistics		57.085	7	.000

Block 1: Method = Enter

Omnibus Tests of Model Coefficients

		Chi-square	df	Sig.
Step 1	Step	56.900	7	.000
	Block	56.900	7	.000
	Model	56.900	7	.000

Model Summary

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	293.056	.173	.251

Classification Table[a]

			Predicted		
			BLESSURE		Percentage Correct
	Observed		.00	1.00	
Step 1	BLESSURE	.00	206	13	94.1
		1.00	54	27	33.3
	Overall Percentage				77.7

a. The cut value is .500

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)	95.0% C.I.for EXP(B) Lower	Upper
Step 1[a]	VORST	.073	.617	.014	1	.905	1.076	.321	3.607
	NEERSLAG	1.287	.337	14.571	1	.000	3.621	1.870	7.010
	NEERSLAG by VORST	1.728	.777	4.951	1	.026	5.631	1.229	25.807
	SPORT			8.471	3	.037			
	SPORT(1)	.600	.397	2.280	1	.131	1.822	.836	3.972
	SPORT(2)	-.603	.444	1.844	1	.174	.547	.229	1.306
	SPORT(3)	-.199	.421	.223	1	.637	.820	.359	1.871
	TIJD	.105	.058	3.223	1	.073	1.110	.990	1.244
	Constant	-3.332	.887	14.107	1	.000	.036		

a. Variable(s) entered on step 1: VORST, NEERSLAG, NEERSLAG * VORST , SPORT, TIJD.

Block 2: Method = Enter

Omnibus Tests of Model Coefficients

		Chi-square	df	Sig.
Step 1	Step	.436	1	.509
	Block	.436	1	.509
	Model	57.336	8	.000

Model Summary

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	292.620	.174	.253

Classification Table[a]

			Predicted		
			BLESSURE		Percentage Correct
	Observed		.00	1.00	
Step 1	BLESSURE	.00	207	12	94.5
		1.00	55	26	32.1
	Overall Percentage				77.7

a. The cut value is .500

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)	95.0% C.I.for EXP(B) Lower	Upper
Step 1[a]	VORST	-.042	.641	.004	1	.948	.959	.273	3.372
	NEERSLAG	1.207	.357	11.405	1	.001	3.343	1.659	6.734
	NEERSLAG by VORST	1.788	.783	5.211	1	.022	5.975	1.288	27.724
	SPORT			8.327	3	.040			
	SPORT(1)	.612	.399	2.355	1	.125	1.844	.844	4.031
	SPORT(2)	-.582	.446	1.701	1	.192	.559	.233	1.340
	SPORT(3)	-.178	.422	.177	1	.674	.837	.366	1.915
	TIJD	.100	.059	2.917	1	.088	1.105	.985	1.240
	KUNSTGRA	.215	.325	.437	1	.509	1.239	.656	2.342
	Constant	-3.344	.891	14.075	1	.000	.035		

a. Variable(s) entered on step 1: KUNSTGRA.

c. (5 punten) Wat is nu je conclusie met betrekking tot het spelen op kunstgras? Is er een relatie met blessures en hoe zit die relatie eruit?

d. (2 punten) Is er een effect van het type sport dat gespeeld wordt? Hoe ziet deze relatie eruit?

e. (3 punten) Hoe kun je de interactie tussen NEERSLAG en VORST interpreteren?

f. (5 punten) Stel dat de data niet bij 300 verschillende velden verzameld is, maar dat de onderzoekers 6 verschillende wedstrijden op verschillende 50 velden onderzocht hebben (ieder veld is op 6 verschillende dagen bekeken dus). Hoe zou de data dan geanalyseerd moeten worden?