

Het gebruik van een grafische rekenmachine is toegestaan tijdens dit tentamen, alsmede één A4-tje met aantekeningen.

1. (a) In de appendix van deze vraag, is een dataset gegeven met de corresponderende boxplot. Bereken de hoogte van de box, alsmede de lengte van de whiskers tot de punten boven resp. onder de box.
- (b) De tabel hieronder geeft de mediaan en de lower- and upper hinges van verscheidene Box-Cox machtstransformaties van de dataset gegeven in de appendix:

Power	-1	-0.5	0	0.5
H_L	-2.6301	-1.8105	-1.2892	-0.9502
Median	-0.3371	-0.3126	-0.2904	-0.2703
H_U	0.3171	0.3473	0.3816	0.4206

Geef de definitie van de familie van Box-Cox machtstransformaties.

Welke van de gegeven transformaties is het meest geschikt om de asymmetrie van de oorspronkelijke dataset te corrigeren?

- (c) In de appendix bij deze vraag is gegeven een scatterplot met een **lowess** curve van een response variabele t.o.v. een verklarende variabele. Om het verband te lineariseren, wil men de familie van Box-Cox machtstransformaties gebruiken. Geef aan welke soort macht gebruikt moet worden om de response en/of de verklarende variabele te transformeren.
2. In een experiment stelt men de waarden van twee variabelen X_1 and X_2 vast en meet men de waarde van een response variabele Y . Dit geeft de volgende tabel:

X_1	2	3	-6	1
X_2	0	-1	-2	3
Y	7	13	-13	1

De corresponderende regressie vergelijking is $Y_i = \alpha + \beta_1 X_{i1} + \beta_2 X_{i2} + \epsilon_i$.

- (a) Bepaal de kleinste kwadraten schatters A , B_1 and B_2 voor α , β_1 en β_2

- (b) Men voert twee extra experimenten uit met de volgende waarden

$$\begin{array}{c|cc} X_1 & -a & a \\ X_2 & 1 & -1 \end{array}$$

voor een $a \in \mathbb{R}$. Verder meet men de overeenkomstige response variabele en men bepaalt de kleinste kwadraten schatters op basis van alle 6 de metingen. Voor welke waarde van $a \in \mathbb{R}$ zijn A , B_1 and B_2 ongecorrleerd?

3. Stel gegeven $(X_1, X_2, X_3) \sim N(\mu, \Sigma)$, met

$$\mu = (0, 1, 1) \text{ and } \Sigma = \begin{pmatrix} 4 & -1 & 0 \\ -1 & 2 & 1 \\ 0 & 1 & 4 \end{pmatrix}.$$

- (a) Bereken $\text{Cov}(X_1 + X_2 + X_3, X_1 - X_2 - X_3)$
 - (b) Bereken $P(X_1 + X_2 - 3X_3 > -3)$
4. In de appendix bij deze vraag is de gewichtstoename van ratten weergegeven bij het volgen van 4 verschillende voedingspatronen; deze onderscheiden zich naar het type proteïne (**source**) en de hoeveelheid proteïne (**amount**). Bovendien zijn gedeelten van de output gegeven, die verkregen is door het fitten van two-way ANOVA-modellen, met en zonder interactie. Het is van belang om te weten of de gewichtstoename verschilt bij het type en de hoeveelheid proteïne.
- (a) Vul de tabel aan die overeenkomt met het model zonder interactie (zie Appendix). Is er een effect voor het type proteïne?
 - (b) Test de nulhypothese dat geen enkele factor van invloed is op de gewichtstoename op significantieniveau 0.05 in het two-way ANOVA model zonder interactie. Geef de waarde van de toetsingsgrootheid en geef uw conclusie aangaande de nulhypothese.
 - (c) Toets of er een interactie effect is op significantieniveau 0.05. Geef de waarde van de toetsingsgrootheid en geef de conclusie aangaande de nulhypothese.
5. (a) Geef de definitie van een Generalised Linear Model
- (b) Formuleer de Gauss-Markov Stelling
- (c) Stel we hebben te maken met het standaard multivariate regressie model, met $Y = X\beta + \epsilon$, met Normaal verdeelde fouten met dezelfde variantie. Schets het bewijs van de stelling dat

$$\frac{\text{RSS}}{\sigma^2} \sim \chi^2(n - p)$$

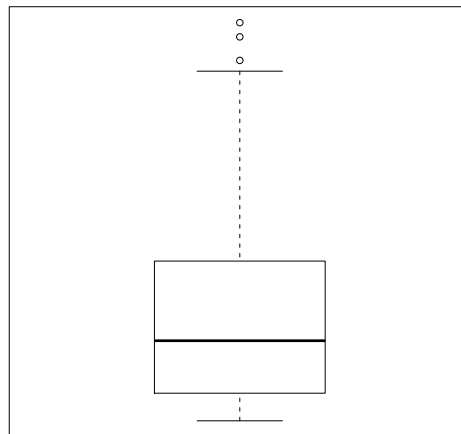
6. (a) Welke diagnostiek(en) heb je tot je beschikking om high leverage points in een dataset te identificeren? Geef de naam van de diagnostiek(en) en de definitie ervan.
- (b) Welke diagnostiek(en) heb je tot je beschikking om punten te identificeren die van gemeenschappelijke invloed op de coëfficiënten zijn? Geef de naam van de diagnostiek(en) en leg uit hoe dat werkt.

Appendix bij Opgave 1.

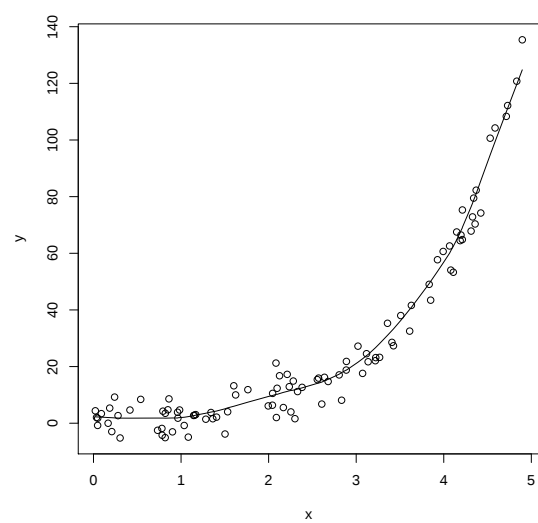
•

0.027	0.032	0.033	0.033	0.038	0.055	0.070	0.076	0.088	0.108
0.114	0.151	0.162	0.167	0.208	0.209	0.227	0.232	0.234	0.237
0.244	0.256	0.257	0.261	0.273	0.278	0.284	0.314	0.329	0.360
0.387	0.387	0.391	0.410	0.414	0.434	0.440	0.450	0.472	0.543
0.569	0.634	0.649	0.657	0.672	0.674	0.721	0.730	0.732	0.738
0.758	0.791	0.820	0.843	0.857	0.869	0.892	0.902	0.902	0.906
0.931	0.934	1.010	1.042	1.044	1.058	1.076	1.166	1.222	1.241
1.275	1.299	1.342	1.420	1.434	1.496	1.516	1.613	1.653	1.718
1.776	1.793	1.809	1.815	1.854	1.885	1.895	1.944	2.117	2.216
2.219	2.535	2.873	2.898	3.147	3.148	3.175	3.272	3.484	3.612

•



•



Appendix bij Opgave 4.¹ Gewichtstoename bij ratten bij het geven van vier verschillende diëten, onderscheiden naar het type proteïne (Beef and Cereal) en de hoeveelheid proteïne (Low and High):

Proteïne type Proteïne hoeveelheid	Beef Low	Beef High	Cereal Low	Cereal High
	90	73	107	98
	76	102	95	74
	90	118	97	56
	64	104	80	111
	86	81	98	95
	51	107	74	88
	72	100	74	82
	90	87	67	77
	95	117	89	86
	78	111	58	92

Output of fitting a two-way ANOVA model without interaction:

Response: weight.gain

Terms added sequentially (first to last)

	Df	Sum of Sq	Mean Sq	F Value	Pr(F)
source ..		220.9			0.3450114
amount ..		1299.6			0.0259568
Residuals ..		8933.0			

Output of fitting a two-way ANOVA model with interaction:

Response: weight.gain

Terms added sequentially (first to last)

	Df	Sum of Sq	Mean Sq	F Value	Pr(F)
source ..		220.9		0.987949	0.3268783
amount ..		1299.6		5.812309	0.0211449
source:amount ..					
Residuals ..		8049.4			

¹Data uit Snedecor, G.W. and Cochran, G.C. (1967) *Statistical Methods* 6th edition, Ames, Iowa: Iowa State University Press, 347.

Antwoorden:

1a hoogte: 1.1895; onderste whisker: 0.2485; bovenste whisker: 1.71.

1b Definitie, zie Fox, pag. 50. Optimale transformatie: $\sqrt{x}$.

1c Transformeer de response variabele met macht $p < 1$ en/of transformeer de verklarende variabele met macht $p > 1$.

2a $A = 2$, $B_1 = 1656/556 = 2.978$, and $B_2 = -784/556 = -1.410$.

2b $a = 6$.

3a -4

3b 0.5675

4a De volledige tabel wordt gegeven door

	Df	Sum of Sq	Mean Sq	F Value	Pr(F)
source	1	220.9	220.900	0.914956	0.3450114
amount	1	1299.6	1299.600	5.382872	0.0259568
Residuals	37	8933.0	241.432		

Nee, geen effect van de soort proteïne.

4b $F = 3.153$. Niet verwerpen.

4c $F = 3.952$; verwerp de nulhypothese niet, er is geen interactie effect.

6a hat-values (zie Fox pag. 244)

6b gebruik Added Variable plots (zie Fox pag. 255 e.v.)

a	0	1	2	3	4	5	6	7	8	9
0.0	5000	4960	4920	4880	4840	4801	4761	4721	4681	4641
0.1	4602	4562	4522	4483	4443	4404	4364	4325	4286	4247
0.2	4207	4168	4129	4090	4052	4013	3974	3936	3897	3859
0.3	3821	3783	3745	3707	3669	3632	3594	3557	3520	3483
0.4	3446	3409	3372	3336	3300	3264	3228	3192	3156	3121
0.5	3085	3050	3015	2981	2946	2912	2877	2843	2810	2776
0.6	2743	2709	2676	2643	2611	2578	2546	2514	2483	2451
0.7	2420	2389	2358	2327	2296	2266	2236	2206	2177	2148
0.8	2119	2090	2061	2033	2005	1977	1949	1922	1894	1867
0.9	1841	1814	1788	1762	1736	1711	1685	1660	1635	1611
1.0	1587	1562	1539	1515	1492	1469	1446	1423	1401	1379
1.1	1357	1335	1314	1292	1271	1251	1230	1210	1190	1170
1.2	1151	1131	1112	1093	1075	1056	1038	1020	1003	0985
1.3	0968	0951	0934	0918	0901	0885	0869	0853	0838	0823
1.4	0808	0793	0778	0764	0749	0735	0721	0708	0694	0681
1.5	0668	0655	0643	0630	0618	0606	0594	0582	0571	0559
1.6	0548	0537	0526	0516	0505	0495	0485	0475	0465	0455
1.7	0446	0436	0427	0418	0409	0401	0392	0384	0375	0367
1.8	0359	0351	0344	0336	0329	0322	0314	0307	0301	0294
1.9	0287	0281	0274	0268	0262	0256	0250	0244	0239	0233
2.0	0228	0222	0217	0212	0207	0202	0197	0192	0188	0183
2.1	0179	0174	0170	0166	0162	0158	0154	0150	0146	0143
2.2	0139	0136	0132	0129	0125	0122	0119	0116	0113	0110
2.3	0107	0104	0102	0099	0096	0094	0091	0089	0087	0084
2.4	0082	0080	0078	0075	0073	0071	0069	0068	0066	0064
2.5	0062	0060	0059	0057	0055	0054	0052	0051	0049	0048
2.6	0047	0045	0044	0043	0041	0040	0039	0038	0037	0036
2.7	0035	0034	0033	0032	0031	0030	0029	0028	0027	0026
2.8	0026	0025	0024	0023	0023	0022	0021	0021	0020	0019
2.9	0019	0018	0018	0017	0016	0016	0015	0015	0014	0014
3.0	0013	0013	0013	0012	0012	0011	0011	0011	0010	0010
3.1	0010	0009	0009	0009	0008	0008	0008	0008	0007	0007
3.2	0007	0007	0006	0006	0006	0006	0006	0005	0005	0005
3.3	0005	0005	0005	0004	0004	0004	0004	0004	0004	0003
3.4	0003	0003	0003	0003	0003	0003	0003	0003	0003	0002

Tabel 1: *Rechteroverschrijdingskans* $1 - \Phi(a) = P(Z \geq a)$
van de $N(0, 1)$ -variabele Z .

	$\longleftarrow \nu_1 \longrightarrow$								
ν_2	1	2	3	4	5	6	7	8	9
1	161.45	199.50	215.71	224.58	230.16	233.99	236.77	238.88	240.54
2	18.51	19.00	19.16	19.25	19.30	19.33	19.35	19.37	19.38
3	10.13	9.55	9.28	9.12	9.01	8.94	8.89	8.85	8.81
4	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	6.00
5	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.77
6	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.10
7	5.59	4.74	4.35	4.12	3.97	3.87	3.79	3.73	3.68
8	5.32	4.46	4.07	3.84	3.69	3.58	3.50	3.44	3.39
9	5.12	4.26	3.86	3.63	3.48	3.37	3.29	3.23	3.18
10	4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07	3.02
11	4.84	3.98	3.59	3.36	3.20	3.09	3.01	2.95	2.90
12	4.75	3.89	3.49	3.26	3.11	3.00	2.91	2.85	2.80
13	4.67	3.81	3.41	3.18	3.03	2.92	2.83	2.77	2.71
14	4.60	3.74	3.34	3.11	2.96	2.85	2.76	2.70	2.65
15	4.54	3.68	3.29	3.06	2.90	2.79	2.71	2.64	2.59
16	4.49	3.63	3.24	3.01	2.85	2.74	2.66	2.59	2.54
17	4.45	3.59	3.20	2.96	2.81	2.70	2.61	2.55	2.49
18	4.41	3.55	3.16	2.93	2.77	2.66	2.58	2.51	2.46
19	4.38	3.52	3.13	2.90	2.74	2.63	2.54	2.48	2.42
20	4.35	3.49	3.10	2.87	2.71	2.60	2.51	2.45	2.39
21	4.32	3.47	3.07	2.84	2.68	2.57	2.49	2.42	2.37
22	4.30	3.44	3.05	2.82	2.66	2.55	2.46	2.40	2.34
23	4.28	3.42	3.03	2.80	2.64	2.53	2.44	2.37	2.32
24	4.26	3.40	3.01	2.78	2.62	2.51	2.42	2.36	2.30
25	4.24	3.39	2.99	2.76	2.60	2.49	2.40	2.34	2.28
26	4.23	3.37	2.98	2.74	2.59	2.47	2.39	2.32	2.27
27	4.21	3.35	2.96	2.73	2.57	2.46	2.37	2.31	2.25
28	4.20	3.34	2.95	2.71	2.56	2.45	2.36	2.29	2.24
29	4.18	3.33	2.93	2.70	2.55	2.43	2.35	2.28	2.22
30	4.17	3.32	2.92	2.69	2.53	2.42	2.33	2.27	2.21
40	4.08	3.23	2.84	2.61	2.45	2.34	2.25	2.18	2.12
50	4.03	3.18	2.79	2.56	2.40	2.29	2.20	2.13	2.07
60	4.00	3.15	2.76	2.53	2.37	2.25	2.17	2.10	2.04
70	3.98	3.13	2.74	2.50	2.35	2.23	2.14	2.07	2.02
80	3.96	3.11	2.72	2.49	2.33	2.21	2.13	2.06	2.00
90	3.95	3.10	2.71	2.47	2.32	2.20	2.11	2.04	1.99
100	3.94	3.09	2.70	2.46	2.31	2.19	2.10	2.03	1.97
110	3.93	3.08	2.69	2.45	2.30	2.18	2.09	2.02	1.97
120	3.92	3.07	2.68	2.45	2.29	2.18	2.09	2.02	1.96

Tabel 2: Right critical values of the $F_\alpha(\nu_1, \nu_2)$ from the $F(\nu_1, \nu_2)$ distribution with $\alpha = 0.05$.

$m \backslash p$	0.1	0.05	0.025	0.01	0.005	0.0025	0.001	0.0005
1	3.078	6.314	12.706	31.821	63.657	127.321	318.309	636.619
2	1.886	2.920	4.303	6.965	9.925	14.089	22.327	31.599
3	1.638	2.353	3.182	4.541	5.841	7.453	10.215	12.924
4	1.533	2.132	2.776	3.747	4.604	5.598	7.173	8.610
5	1.476	2.015	2.571	3.365	4.032	4.773	5.893	6.869
6	1.440	1.943	2.447	3.143	3.707	4.317	5.208	5.959
7	1.415	1.895	2.365	2.998	3.499	4.029	4.785	5.408
8	1.397	1.860	2.306	2.896	3.355	3.833	4.501	5.041
9	1.383	1.833	2.262	2.821	3.250	3.690	4.297	4.781
10	1.372	1.812	2.228	2.764	3.169	3.581	4.144	4.587
11	1.363	1.796	2.201	2.718	3.106	3.497	4.025	4.437
12	1.356	1.782	2.179	2.681	3.055	3.428	3.930	4.318
13	1.350	1.771	2.160	2.650	3.012	3.372	3.852	4.221
14	1.345	1.761	2.145	2.624	2.977	3.326	3.787	4.140
15	1.341	1.753	2.131	2.602	2.947	3.286	3.733	4.073
16	1.337	1.746	2.120	2.583	2.921	3.252	3.686	4.015
17	1.333	1.740	2.110	2.567	2.898	3.222	3.646	3.965
18	1.330	1.734	2.101	2.552	2.878	3.197	3.610	3.922
19	1.328	1.729	2.093	2.539	2.861	3.174	3.579	3.883
20	1.325	1.725	2.086	2.528	2.845	3.153	3.552	3.850
21	1.323	1.721	2.080	2.518	2.831	3.135	3.527	3.819
22	1.321	1.717	2.074	2.508	2.819	3.119	3.505	3.792
23	1.319	1.714	2.069	2.500	2.807	3.104	3.485	3.768
24	1.318	1.711	2.064	2.492	2.797	3.091	3.467	3.745
25	1.316	1.708	2.060	2.485	2.787	3.078	3.450	3.725
26	1.315	1.706	2.056	2.479	2.779	3.067	3.435	3.707
27	1.314	1.703	2.052	2.473	2.771	3.057	3.421	3.690
28	1.313	1.701	2.048	2.467	2.763	3.047	3.408	3.674
29	1.311	1.699	2.045	2.462	2.756	3.038	3.396	3.659
30	1.310	1.697	2.042	2.457	2.750	3.030	3.385	3.646
40	1.303	1.684	2.021	2.423	2.704	2.971	3.307	3.551
50	1.299	1.676	2.009	2.403	2.678	2.937	3.261	3.496
∞	1.282	1.645	1.960	2.326	2.576	2.807	3.090	3.291

Tabel 3: Right critical values $t_{m,p}$ of the t -distribution with m degrees of freedom corresponding to right tail probability p : $P(T_m \geq t_{m,p}) = p$. The last row in the table are right critical values of the $N(0,1)$ distribution: $t_{\infty,p} = z_p$