VRIJE UNIVERSITEIT AMSTERDAM

Exam Statistical Models on 14-12-2009

You may only use a calculator. The answers to the exercises need to be unambiguous ('no double answers') and clearly, but preferably concisely (Ned: 'bondig'), motivated. Items of exercise 1 are awarded with max. **2** points. Items of exercises 2,3 and 4 are awarded with max. **3** points (max. 42 point in total). Exam score is min(10,(total score)/4). Final mark is computed as denoted on the web site. **All statistical tests are to be performed using $\alpha = 0.05$, unless stated otherwise. Critical values of the F-distribution are displayed at the end.**

1. We study the yield (Ned: 'opbrengst') from a certain chemical process, depending on two input factors: mixing speed (S: 60, 70 ,80) and formulation (F: 1, 2). The data is displayed in the Table below.

	60	70	80
	189.7	185.1	189.0
1	188.6	179.4	193.0
	190.1	177.3	191.1
	165.1	161.7	163.3
2	165.9	159.8	166.6
	167.6	161.6	170.3

Moreover, we have the following sums of squares: SS(F) = 2253.44, SS(S) = 230.81, SS(F*S) = 18.58 and SSE = 71.87.

(a) Formulate an appropriate model plus all assumptions and constraints.

(b) Does yield depend on either speed or formulation?

(c) Does their appear to be an interaction effect between the two factors?

(d) Calculate estimates of the mean effects.

(e) We could also have modeled Speed as a continuous variable in a linear regression setting. What are the advantages/disadvantages with respect to the model you formulated under a.?

(f) Given the data, do think it is wise to use Speed as linear predictor, as suggested in the previous exercise?

2. Figure 1 shows two series of measurements. Series 1, displayed with a $\circ$, was obtained under colder circumstances than Series 2, displayed by $\times$. For each series the data can be described by a nonlinear model of the following type

$$Y_i = \theta_1 + \frac{\theta_2 x_i}{\theta_3 + x_i} + \varepsilon_i,$$

where $\varepsilon_1, \ldots, \varepsilon_{20}$ are independent measurement errors which are assumed to be normally distributed with mean 0 and variance σ^2. The values of $\theta_1, \theta_2, \theta_3$ and σ^2 are unknown. It is suspected that the value of θ_2 is influenced by the temperature and that the other parameter values are insensitive to temperature.

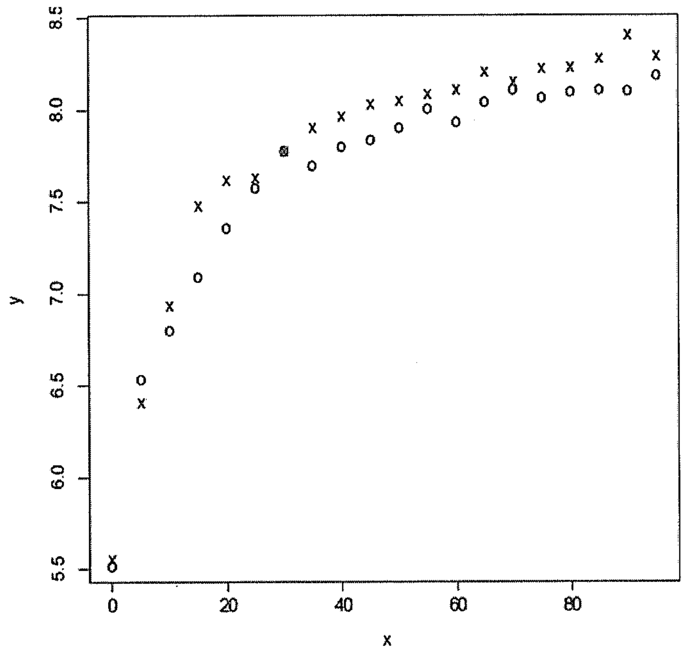

Figure 1: Two series of measurements (Series 1: $\circ$, Series 2: $\times$)

(a) How is the least squares estimator $\hat{\theta}$ of $\theta = (\theta_1, \theta_2, \theta_3)^T$ defined?

(b) Find suitable starting values for estimating $\theta = (\theta_1, \theta_2, \theta_3)^T$ for Series 1.

(c) The table below lists the parameter estimations with their 95% confidence intervals for the two series. What do you conclude, based on these results, with respect to the suspicion about $\theta_1, \theta_2, \theta_3$ being (in)sensitive to temperature?

	Series 1		Series 2	
	estimate	interval	estimate	interval
θ_1	5.54	(5.414,5.657)	5.52	(5.383,5.648)
θ_2	2.94	(2.813,3.068)	3.11	(2.967,3.244)
θ_3	11.96	(10.016,13.899)	10.96	(9.154,12.762)

(d) Describe how the null hypothesis that the values of θ_2 are equal for the two series, can be tested.

3. The negative binomial distribution describes the number of experiments k need before r successes are observed, where each experiment has success probability p. Here, r is supposed to be fixed. It generalizes the geometric distribution, for which $r = 1$. The probability density function of the negative binomial distribution is

$$\mathrm{P}(Y = y) = \binom{y + r - 1}{r - 1} p^r (1 - p)^y.$$

(a) The general form of the exponential family is $f(y, \theta) = \exp(\frac{y\theta - b(\theta)}{\phi/A} + c(y, \phi/A))$. Show that the negative binomial distribution belongs to the exponential family.

(b) Show that the mean $E[Y] = rp/(1-p)$.

(c) Suppose we have covariates to model the linear predictor (η). Describe the steps for estimating the unknown parameters in the linear predictor.

4. Consider the following ARMA(1,1) process.

$$X_t = \alpha X_{t-1} + \beta Z_{t-1} + Z_t$$

where $Z_i \sim N(0, \sigma^2)$.

(a) We assume a stationary series. What does this mean?

(b) Estimate α and β using the Yule-Walker equations and the following estimates of the autocorrelations. $\hat{\gamma}_X(1) = 0.7, \hat{\gamma}_X(2) = -0.3$.

(c) Now suppose there is a trend and seasonality present in the data. Discuss the small trend method and explain how to use it for removing those structural components.

Table V Percentage Points of the F-Distribution (*continued*)

$f_{0.05,\nu_1,\nu_2}$

ν_2 \ ν_1	Degrees of freedom for the numerator (ν_1) 1	2	3	4	5	6	7	8	9	10	12	15	20	24	30	40	60	120	∞
1	161.4	199.5	215.7	224.6	230.2	234.0	236.8	238.9	240.5	241.9	243.9	245.9	248.0	249.1	250.1	251.1	252.2	253.3	254.3
2	18.51	19.00	19.16	19.25	19.30	19.33	19.35	19.37	19.38	19.40	19.41	19.43	19.45	19.45	19.46	19.47	19.48	19.49	19.50
3	10.13	9.55	9.28	9.12	9.01	8.94	8.89	8.85	8.81	8.79	8.74	8.70	8.66	8.64	8.62	8.59	8.57	8.55	8.53
4	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	6.00	5.96	5.91	5.86	5.80	5.77	5.75	5.72	5.69	5.66	5.63
5	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.77	4.74	4.68	4.62	4.56	4.53	4.50	4.46	4.43	4.40	4.36
6	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.10	4.06	4.00	3.94	3.87	3.84	3.81	3.77	3.74	3.70	3.67
7	5.59	4.74	4.35	4.12	3.97	3.87	3.79	3.73	3.68	3.64	3.57	3.51	3.44	3.41	3.38	3.34	3.30	3.27	3.23
8	5.32	4.46	4.07	3.84	3.69	3.58	3.50	3.44	3.39	3.35	3.28	3.22	3.15	3.12	3.08	3.04	3.01	2.97	2.93
9	5.12	4.26	3.86	3.63	3.48	3.37	3.29	3.23	3.18	3.14	3.07	3.01	2.94	2.90	2.86	2.83	2.79	2.75	2.71
10	4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07	3.02	2.98	2.91	2.85	2.77	2.74	2.70	2.66	2.62	2.58	2.54
11	4.84	3.98	3.59	3.36	3.20	3.09	3.01	2.95	2.90	2.85	2.79	2.72	2.65	2.61	2.57	2.53	2.49	2.45	2.40
12	4.75	3.89	3.49	3.26	3.11	3.00	2.91	2.85	2.80	2.75	2.69	2.62	2.54	2.51	2.47	2.43	2.38	2.34	2.30
13	4.67	3.81	3.41	3.18	3.03	2.92	2.83	2.77	2.71	2.67	2.60	2.53	2.46	2.42	2.38	2.34	2.30	2.25	2.21
14	4.60	3.74	3.34	3.11	2.96	2.85	2.76	2.70	2.65	2.60	2.53	2.46	2.39	2.35	2.31	2.27	2.22	2.18	2.13
15	4.54	3.68	3.29	3.06	2.90	2.79	2.71	2.64	2.59	2.54	2.48	2.40	2.33	2.29	2.25	2.20	2.16	2.11	2.07
16	4.49	3.63	3.24	3.01	2.85	2.74	2.66	2.59	2.54	2.49	2.42	2.35	2.28	2.24	2.19	2.15	2.11	2.06	2.01
17	4.45	3.59	3.20	2.96	2.81	2.70	2.61	2.55	2.49	2.45	2.38	2.31	2.23	2.19	2.15	2.10	2.06	2.01	1.96
18	4.41	3.55	3.16	2.93	2.77	2.66	2.58	2.51	2.46	2.41	2.34	2.27	2.19	2.15	2.11	2.06	2.02	1.97	1.92
19	4.38	3.52	3.13	2.90	2.74	2.63	2.54	2.48	2.42	2.38	2.31	2.23	2.16	2.11	2.07	2.03	1.98	1.93	1.88
20	4.35	3.49	3.10	2.87	2.71	2.60	2.51	2.45	2.39	2.35	2.28	2.20	2.12	2.08	2.04	1.99	1.95	1.90	1.84
21	4.32	3.47	3.07	2.84	2.68	2.57	2.49	2.42	2.37	2.32	2.25	2.18	2.10	2.05	2.01	1.96	1.92	1.87	1.81
22	4.30	3.44	3.05	2.82	2.66	2.55	2.46	2.40	2.34	2.30	2.23	2.15	2.07	2.03	1.98	1.94	1.89	1.84	1.78
23	4.28	3.42	3.03	2.80	2.64	2.53	2.44	2.37	2.32	2.27	2.20	2.13	2.05	2.01	1.96	1.91	1.86	1.81	1.76
24	4.26	3.40	3.01	2.78	2.62	2.51	2.42	2.36	2.30	2.25	2.18	2.11	2.03	1.98	1.94	1.89	1.84	1.79	1.73
25	4.24	3.39	2.99	2.76	2.60	2.49	2.40	2.34	2.28	2.24	2.16	2.09	2.01	1.96	1.92	1.87	1.82	1.77	1.71
26	4.23	3.37	2.98	2.74	2.59	2.47	2.39	2.32	2.27	2.22	2.15	2.07	1.99	1.95	1.90	1.85	1.80	1.75	1.69
27	4.21	3.35	2.96	2.73	2.57	2.46	2.37	2.31	2.25	2.20	2.13	2.06	1.97	1.93	1.88	1.84	1.79	1.73	1.67
28	4.20	3.34	2.95	2.71	2.56	2.45	2.36	2.29	2.24	2.19	2.12	2.04	1.96	1.91	1.87	1.82	1.77	1.71	1.65
29	4.18	3.33	2.93	2.70	2.55	2.43	2.35	2.28	2.22	2.18	2.10	2.03	1.94	1.90	1.85	1.81	1.75	1.70	1.64
30	4.17	3.32	2.92	2.69	2.53	2.42	2.33	2.27	2.21	2.16	2.09	2.01	1.93	1.89	1.84	1.79	1.74	1.68	1.62
40	4.08	3.23	2.84	2.61	2.45	2.34	2.25	2.18	2.12	2.08	2.00	1.92	1.84	1.79	1.74	1.69	1.64	1.58	1.51
60	4.00	3.15	2.76	2.53	2.37	2.25	2.17	2.10	2.04	1.99	1.92	1.84	1.75	1.70	1.65	1.59	1.53	1.47	1.39
120	3.92	3.07	2.68	2.45	2.29	2.17	2.09	2.02	1.96	1.91	1.83	1.75	1.66	1.61	1.55	1.55	1.43	1.35	1.25
∞	3.84	3.00	2.60	2.37	2.21	2.10	2.01	1.94	1.88	1.83	1.75	1.67	1.57	1.52	1.46	1.39	1.32	1.22	1.00

Degrees of freedom for the denominator (ν_2)

659

Figure 2: Right-critical values at $\alpha = 0.05$ for the F_{ν_1,ν_2} distributions