UNIVERSITY OF OSLO

Faculty of Mathematics and Natural Sciences

Exam in GEO4310 – Stochastic methods in hydrology
Day of exam: 5 December 2016
Exam hours: 14:30 – 17:30
This examination paper consists of 4 pages (plus an appendix).
Appendices: Yes
Permitted materials: Calculator

Make sure that your copy of this examination paper is complete before answering.

Task 1 Definitions (10 points):

Give a brief definition or summary for each of the following:

a) Colinearity
b) Stochastic vs. Determinstic
c) Descriptors of Random Variables (Three categories, and parameters therein)
d) Box-Jenkins or Delta Approach
e) Stationarity

Task 2 About goodness-of-fit method (~30 points)

In order to calculate the design flood, you need to know which probability distribution fits your data best. Based on a 35-year record of yearly maximum discharge data from a river station, the following statistics are calculated (with n = 35).

Average: $\overline{Y}$ =5.257 ; where Y = lnQ (natural logarithm).
Standard deviation: $S_Y = 0.686$

35 years data are classified into 5 classes as follows:

Class	Observed number
$Q \leq 100$	5
$100 < Q \leq 150$	9
$150 < Q \leq 200$	7
$200 < Q \leq 300$	6
$Q \geq 300$	8

Use Chi-square method (with significance level **α = 5%**) to test if the data are **<u>log-normally</u>** distributed.

Task 3 Regression analysis (~30 points)

In multiple regression the general equation is given as:

$$Y = a + b_1X1 + b_2X2 + \cdots, +b_kXk + e$$

Where a and b1 – bk are regression coefficients, e is a regression error term, X are the independent variables and Y is the dependent variable.

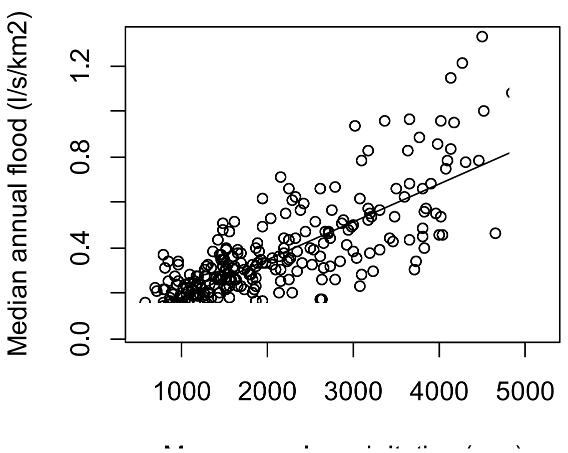
a) What are the four basic assumption for standard linear multiple regression analysis?

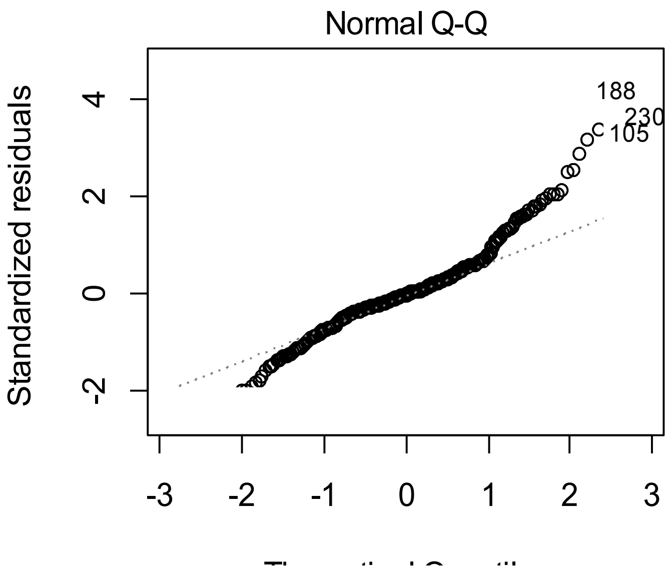
b) The explained variance is used for evaluating the performance of a regression model.

   i) Describe how explained variance is calculated

   ii) What is the explained variance for a perfect model and for the worst model?

c) Below is a plot of Median annual flood (dependent variable) as a function of annual precipitation (independent variable). Which assumptions are not fulfilled in this case





Below is the output from a multiple regression analysis using four covariates. Which of the regression coefficients are significant?

	Estimate	**t-value**	**p-value**
Intercept	-1.94	-39.9	<0.0000001
Annual precipitation (mm)	0.00042	21.99	<0.0000001
Catchment area (km2)	-0.0001	-6.31	<0.0000001
Effective lake percentage	-0.032	6.75	<0.0000001
Catchment gradient (m/km)	-0.000069	1.73	0.0845

Task 4 Probability (~30 points)

A) The values of annual precipitation at College Station, Texas, from 1911 to 1979 are provided below (in inches!). What is the probability that the annual precipitation, P, in any year will be less than 35 in.? Greater than 45 in.? Between 35 and 45 in?

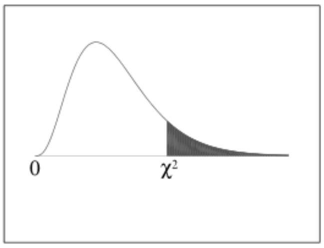
B) Assuming that precipitation is an independent process, calculate the probabilitiy that there will be two successive years of precipitation less than 35. in. Compare this estimated value with the data in the table.

C) Is this a valid assumption? Explain.

Year	1910	1920	1930	1940	1950	1960	1970
0		48.7	44.8	49.3	31.2	46.0	33.9
1	39.9	44.1	34.0	44.2	27.0	44.3	31.7
2	31.0	42.8	45.6	41.7	37.0	37.8	31.5
3	42.3	48.4	37.3	30.8	46.8	29.6	59.6
4	42.1	34.2	43.7	53.6	26.9	35.1	50.5
5	41.1	32.4	41.8	34.5	25.4	49.7	38.6
6	28.7	46.4	41.1	50.3	23.0	36.6	43.4
7	16.8	38.9	31.2	43.8	56.5	32.5	28.7
8	34.1	37.3	35.2	21.6	43.4	61.7	32.0
9	56.4	50.6	35.1	47.1	41.3	47.4	51.8

Appendices

Chi-Square Distribution Table



The shaded area is equal to α for $\chi^2 = \chi^2_\alpha$.

df	$\chi^2_{.995}$	$\chi^2_{.990}$	$\chi^2_{.975}$	$\chi^2_{.950}$	$\chi^2_{.900}$	$\chi^2_{.100}$	$\chi^2_{.050}$	$\chi^2_{.025}$	$\chi^2_{.010}$	$\chi^2_{.005}$
1	0.000	0.000	0.001	0.004	0.016	2.706	3.841	5.024	6.635	7.879
2	0.010	0.020	0.051	0.103	0.211	4.605	5.991	7.378	9.210	10.597
3	0.072	0.115	0.216	0.352	0.584	6.251	7.815	9.348	11.345	12.838
4	0.207	0.297	0.484	0.711	1.064	7.779	9.488	11.143	13.277	14.860
5	0.412	0.554	0.831	1.145	1.610	9.236	11.070	12.833	15.086	16.750
6	0.676	0.872	1.237	1.635	2.204	10.645	12.592	14.449	16.812	18.548
7	0.989	1.239	1.690	2.167	2.833	12.017	14.067	16.013	18.475	20.278
8	1.344	1.646	2.180	2.733	3.490	13.362	15.507	17.535	20.090	21.955
9	1.735	2.088	2.700	3.325	4.168	14.684	16.919	19.023	21.666	23.589
10	2.156	2.558	3.247	3.940	4.865	15.987	18.307	20.483	23.209	25.188
11	2.603	3.053	3.816	4.575	5.578	17.275	19.675	21.920	24.725	26.757
12	3.074	3.571	4.404	5.226	6.304	18.549	21.026	23.337	26.217	28.300
13	3.565	4.107	5.009	5.892	7.042	19.812	22.362	24.736	27.688	29.819
14	4.075	4.660	5.629	6.571	7.790	21.064	23.685	26.119	29.141	31.319
15	4.601	5.229	6.262	7.261	8.547	22.307	24.996	27.488	30.578	32.801
16	5.142	5.812	6.908	7.962	9.312	23.542	26.296	28.845	32.000	34.267
17	5.697	6.408	7.564	8.672	10.085	24.769	27.587	30.191	33.409	35.718
18	6.265	7.015	8.231	9.390	10.865	25.989	28.869	31.526	34.805	37.156
19	6.844	7.633	8.907	10.117	11.651	27.204	30.144	32.852	36.191	38.582
20	7.434	8.260	9.591	10.851	12.443	28.412	31.410	34.170	37.566	39.997
21	8.034	8.897	10.283	11.591	13.240	29.615	32.671	35.479	38.932	41.401
22	8.643	9.542	10.982	12.338	14.041	30.813	33.924	36.781	40.289	42.796
23	9.260	10.196	11.689	13.091	14.848	32.007	35.172	38.076	41.638	44.181
24	9.886	10.856	12.401	13.848	15.659	33.196	36.415	39.364	42.980	45.559
25	10.520	11.524	13.120	14.611	16.473	34.382	37.652	40.646	44.314	46.928
26	11.160	12.198	13.844	15.379	17.292	35.563	38.885	41.923	45.642	48.290
27	11.808	12.879	14.573	16.151	18.114	36.741	40.113	43.195	46.963	49.645
28	12.461	13.565	15.308	16.928	18.939	37.916	41.337	44.461	48.278	50.993
29	13.121	14.256	16.047	17.708	19.768	39.087	42.557	45.722	49.588	52.336
30	13.787	14.953	16.791	18.493	20.599	40.256	43.773	46.979	50.892	53.672
40	20.707	22.164	24.433	26.509	29.051	51.805	55.758	59.342	63.691	66.766
50	27.991	29.707	32.357	34.764	37.689	63.167	67.505	71.420	76.154	79.490
60	35.534	37.485	40.482	43.188	46.459	74.397	79.082	83.298	88.379	91.952
70	43.275	45.442	48.758	51.739	55.329	85.527	90.531	95.023	100.425	104.215
80	51.172	53.540	57.153	60.391	64.278	96.578	101.879	106.629	112.329	116.321
90	59.196	61.754	65.647	69.126	73.291	107.565	113.145	118.136	124.116	128.299
100	67.328	70.065	74.222	77.929	82.358	118.498	124.342	129.561	135.807	140.169

Formulary

Statistics

Sample mean:

$$\hat{\mu} = m_X = \frac{1}{N}\sum_{i=1}^{N} x_i$$

Sample variance:

$$\hat{\sigma}^2 = s_X{}^2 = \frac{1}{N-1}\sum_{i=1}^{N}\left(x_i - m_X\right)^2$$

Sample coefficient of variation:

$$C\hat{V}_X = \frac{s_X}{m_X}$$

Sample covariance:

$$\hat{\rho}_{X,Y} = r_{X,Y} = \frac{\sum_{i=1}^{N}\left(x_i - m_X\right)\cdot\left(y_i - m_Y\right)}{\left(N-1\right)\cdot s_X \cdot s_Y}$$

Cumulative probability function:

$$P_X\left(x_i\right) \equiv \Pr\left\{X \le x_i\right\} = \sum_{X \le x_i} p_x\left(x_i\right)$$

Weibull plotting-position:

$$q = \frac{i}{N+1}$$

Probability theory

The *conditional probability* of *A* to occur, given that *B* occurs

$$P(A \mid B) = P(A \cap B) / P(B)$$

where:

P(A|B) = the (conditional) probability that event A will occur given that event B has occurred already

$P(A \cap B)$ = the (unconditional) probability that event A and event B occur

The probability that A or B occurs, in case of dependent events:

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

$P(A \cup B) = P(A)+P(B)- P(A \mid B)\cdot P(B)$ if A and B are dependent

$P(A \cup B) = P(A)+P(B)- P(A)\cdot P(B)$ if A and B are independent

$P(A \cup B) = P(A)+P(B)$ if A and B are mutually exclusive

The probability of both A and B to occur:

$P(A \cap B) = P(A|B)\cdot P(B)$ or $P(A \cap B) = P(B|A)\cdot P(A)$

$P(A \cap B) = P(A)\cdot P(B)$ if A and B are independent

Probability of 'event A does not occur'

$P(A') = 1-P(A)$ probability of A not to occur

$P(A \cap A') = 0$ probability of A occur ***and*** A not occur = 0

$P(A \cup A') = P(A)+P(A')=1$ probability of A occur ***or*** A not occur = 1

Sampling and counting

- In Ordered without replacement case, the first item has *n* ways of selection, and second has *n-1* ways, …. Thus *r* ordered items can be selected from *n* without replacement in:
  $(n)_r = n(n-1)(n-2)\ldots(n-r+1) = n!/(n-r)!$

  (The is called the number of **Permutations** of n items taken r at a time)

- Unordered without replacement case, the number of ways in selecting r items from n items is
  $$\binom{n}{r} = \frac{(n)_r}{r!} = \frac{n!}{(n-r)!r!}$$

(Combinations are arrangements of elements without regard to their order or position).

- In unordered with replacement case, the number of ways in selecting r items from n items is

$$\binom{n+r-1}{r} = \frac{(n+r-1)!}{(n-1)!r!}$$

Probability distribution functions

For a continuous random variable X, if the function f(x) satisfies:

a. f(x)>=0 for all x_i
b. $\int f(x)$ dx = 1 for the whole range of x

then f(x) is a probability density function.

Relationship between f(x) and F(x) is

F(x) = ∫f(x)dx
And f(x) = dF(x)/dx

Since F(x) = P(X ≤ x) it follows that

$\int_a^b f(x)dx$ = P(a < X < b) = F (b) - F (a)

Chi-Square distribution:

$$\chi_c^2 = \sum_{j=1}^{n} \frac{(O_j - E_j)^2}{E_j}$$

Hypergeometric distribution: The probability of getting x successes in a sample of size n drawing from population of size N contains k successes is

$$f_x(x; N, n, k) = \binom{k}{x}\binom{N-k}{n-x} \Big/ \binom{N}{n}$$

$$E(x) = \frac{n \cdot k}{N}$$

$$Var(x) = \frac{n \cdot k(N-k)(N-n)}{N^2(N-1)}$$

Binomial distribution:

$$f_x(x; n, p) = \binom{n}{x} p^x q^{n-x}$$

where x- number of success, n – total number of trials, p – probability of one success, q – probability of one failure

$$E(x) = np$$

$$Var(x) = npq$$

Geometric distribution:

$$f_x(x; p) = pq^{x-1}$$

$$E(x)=1/p$$

$$Var(x)=q/p2$$

Flood and extreme value analysis

Return period:

$$TR_X(x) = \frac{1}{1 - F_X(x)}$$

Gumbel distribution:

$$\Pr\{X \le x\} = F_X(x) = \exp\left[-\exp\left(-\left(\frac{x-\xi}{\alpha}\right)\right)\right]$$

$$\mu_x = \xi + 0.5772\alpha$$
$$\sigma_x^2 = 1.645\alpha^2$$

Usage of the frequency factor, *K*, for the determination of values with return period, *TR*:

for the Gumbel distribution:
K(10) = 1.30
K(100) = 3.14
K(1000) = 5.00