Exercise 3.a - Frequency Analysis

Expected time:	2 hours
Objectives:	After this exercise you will be able to: - calculate the return period and the exceeding probability of an event - calculate the extreme value distribution by Gumbel method - determine intensity-duration-frequency relationships - plot and interpret a Gutenberg-Richter graph

Floods

Return period/exceeding probability

Percentage probability of the *N*-year flood occurring in a particular period

Number of years in period	*N* = Return period in years							
	5	10	20	50	100	200	500	1000
1	20	10	5	2	1	–	–	–
2	36	19	10	4	2	1	–	–
5	67	41	23	10	5	2	1	–
10	89	65	40	18	10	5	2	1
20	99	88	64	33	26	10	4	2
50	–	99	92	64	39	22	9	5
100	–	–	99	87	63	39	18	10
200	–	–	–	98	87	63	33	18
500	–	–	–	–	99	92	63	39
1000	–	–	–	–	–	99	86	63

Exercise

1. What can you say about the probability of the occurrence of Q50 during the next 100 years and during the next 5 years?
2. What about the probability of the occurrence of Q500 during the next 2 years?

   Calculate this probability with the following formula: $R = 1 - \left(1 - \frac{1}{T}\right)^n$

Extreme value distribution by Gumbel method

In this exercise a Gumbel distribution is made. With the help of the outcome, different return periods of rainfall/discharge can be determined. First, an example is given in order to show the procedure. This is followed by an exercise.

Example

For this example the yearly maximum floods of the Clearwater River in Idaho USA are selected (1911-1948) N=38 yrs.

1. Rank the yearly maximum flood values from low to high. So assign lowest rank 1 to the lowest data value and assign the highest rank N to the highest data value. Some authors apply a ranking from high (rank=1) to low (rank=N).

2. Calculate for each observation the left sided probabilities by:

$$P_L = \frac{R}{N+1} \quad \text{Eq 1}$$

Where:
P_L= left sided probability (probability of having less values in the series)
R= is the rank
N= number of observations

3. Determine the return period for each observation.

$$T = \frac{1}{P_R} = \frac{1}{1-P_L} \quad \text{Eq 2}$$

4. Determine the plotting position for each observation.

$$y = -\ln(-\ln P_L) \quad \text{Eq 3}$$

An exercise following this procedure can be solved with a spreadsheet as follows:

Table 1: Original data

Date	Discharge	Date	Discharge	Date	Discharge	Date	Discharge
6-May-11	34600	23-May-19	52000	7-May-31	40800	17-May-39	36400
17-May-11	29400	18-May-20	43600	14-May-31	36500	12-May-40	37100
4-Jun-11	35900	16-Jun-20	42900	16-May-31	36500	25-May-40	29600
13-Jun-11	39500	23-Apr-21	35200	14-Apr-32	28500	13-May-41	28900
21-May-12	55200	20-May-21	69700	14-May-32	72100	14-Apr-42	28900
20-May-12	61900	19-May-22	60600	21-May-32	62200	21-Apr-42	28900
21-Jun-12	38000	26-May-22	52100	13-Jun-32	35100	26-May-42	37100
20-Apr-13	29400	6-Jun-22	62400	15-Jun-32	35100	20-Apr-43	43200
27-Apr-13	30700	8-May-23	38800	27-Apr-33	35800	1-May-43	29600
11-May-13	45800	10-May-23	38800	4-Jun-33	71200	29-May-43	52200
26-May-13	76600	26-May-23	49600	10-Jun-33	81400	11-Jun-43	37100
18-May-14	42200	12-Jun-23	43200	23-Dec-33	43600	19-Jun-43	43200
23-May-14	41500	4-May-24	45600	30-Mar-34	32300	22-Jun-43	40100
3-Jun-14	30700	13-May-24	58900	14-Apr-34	37800	16-May-44	34200
19-May-15	28200	17-Apr-25	41800	25-Apr-34	45900	6-May-45	44400
28-Apr-16	30000	7-May-25	44800	8-May-34	34300	31-May-45	38400
7-May-16	44400	20-May-25	59800	24-May-35	44000	20-Apr-46	33300
5-Jun-16	36600	19-Apr-26	35900	31-May-35	34400	26-Apr-46	33700
9-Jun-16	36600	1-May-26	35900	6-Jun-35	29900	6-May-46	36600
19-Jun-16	56000	21-May-26	32400	19-Apr-36	50600	19-May-46	30000
29-Jun-16	36600	28-Apr-27	46400	5-May-36	49800	28-May-46	36100
15-May-17	63600	17-May-27	64200	15-May-36	63200	4-Jun-46	28300
30-May-17	69700	8-Jun-27	68600	28-May-36	34300	15-Dec-46	33900
9-Jun-17	56800	5-Nov-27	43900	1-Jun-36	32900	8-May-47	69900
17-Jun-17	70500	26-Nov-27	29200	19-May-37	34300	27-May-47	37600
29-Dec-17	37300	9-May-28	65700	28-May-37	32200	9-Jun-47	31200
30-Dec-17	37300	26-May-28	72100	19-Apr-38	63400	18-Apr-48	29400
5-May-18	52800	24-May-29	52700	1-May-38	39400	22-Apr-48	32600
15-May-18	35200	1-Jun-29	28500	17-May-38	31500	8-May-48	33800
10-Jun-18	52800	9-Jun-29	35800	28-May-38	60800	22-May-48	86500
29-Apr-19	30700	25-Apr-30	31000	4-May-39	46400	29-May-48	99000
						22-Jun-48	29600

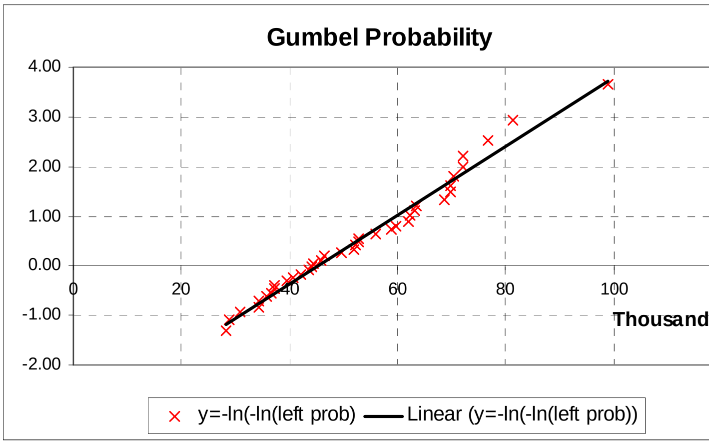
Process

In total 38 years of data. The maximum discharge per year is extracted from the original data and then the steps explained above can be done in a spreadsheet.

Table 2: Sorted values and Gumbel plot

Year	Year Max		Sorted	Rank	Left Prob	TR	y=-ln(-ln(left prob))
1911	39500		28200	1	0.03	1.03	-1.30
1912	61900		28900	2	0.05	1.05	-1.09
1913	76600		31000	3	0.08	1.08	-0.94
1914	42200		34200	4	0.10	1.11	-0.82
1915	28200		34300	5	0.13	1.15	-0.72
1916	56000		35900	6	0.15	1.18	-0.63
1917	70500		36600	7	0.18	1.22	-0.54
1918	52800		37100	8	0.21	1.26	-0.46
1919	52000		37100	9	0.23	1.30	-0.38
1920	43600		39500	10	0.26	1.34	-0.31
1921	69700		40800	11	0.28	1.39	-0.24
1922	62400		42200	12	0.31	1.44	-0.16
1923	49600		43600	13	0.33	1.50	-0.09
1924	58900		44000	14	0.36	1.56	-0.02
1925	59800		44400	15	0.38	1.63	0.05
1926	35900		45900	16	0.41	1.70	0.12
1927	68600		46400	17	0.44	1.77	0.19
1928	72100		49600	18	0.46	1.86	0.26
1929	52700		52000	19	0.49	1.95	0.33
1930	31000		52200	20	0.51	2.05	0.40
1931	40800		52700	21	0.54	2.17	0.48
1932	72100		52800	22	0.56	2.29	0.56
1933	81400		56000	23	0.59	2.44	0.64
1934	45900		58900	24	0.62	2.60	0.72
1935	44000		59800	25	0.64	2.79	0.81
1936	63200		61900	26	0.67	3.00	0.90
1937	34300		62400	27	0.69	3.25	1.00
1938	63400		63200	28	0.72	3.55	1.10
1939	46400		63400	29	0.74	3.90	1.22
1940	37100		68600	30	0.77	4.33	1.34
1941	28900		69700	31	0.79	4.88	1.47
1942	37100		69900	32	0.82	5.57	1.62
1943	52200		70500	33	0.85	6.50	1.79
1944	34200		72100	34	0.87	7.80	1.99
1945	44400		72100	35	0.90	9.75	2.22
1946	36600		76600	36	0.92	13.00	2.53
1947	69900		81400	37	0.95	19.50	2.94
1948	99000		99000	38	0.97	39.00	3.65

5. Using the graphic functionalities of Excel the probability graph can be constructed.



Plot=y	Prob	TR
4	0.981851073	55.09968
3.8	0.977877598	45.20305
3.6	0.973046194	37.10051
3.4	0.967177474	30.46688
3.2	0.960057401	25.03593
3	0.951431993	20.58969
2.8	0.941001954	16.94971
2.6	0.928417664	13.96993
2.4	0.913275261	11.53074
2.2	0.895114927	9.534245
2	0.873423018	7.900331
1.8	0.847640317	6.563416
1.6	0.817179487	5.469846
1.4	0.781455585	4.575729
1.2	0.739934055	3.845179
1	0.692200628	3.24887
0.8	0.638056167	2.76286
0.6	0.577635844	2.367625
0.4	0.511544834	2.047271
0.2	0.440991026	1.78888
0	0.367879441	1.581977
-0.2	0.294816321	1.41807
-0.4	0.224961794	1.290259
-0.6	0.161682814	1.192866
-0.8	0.108008978	1.121088
-1	0.065988036	1.07065

Finally since the ‘y’ axis is directly related to the probability or the return period, the values are exchangeable. The next step is to replace de values in the ‘y; axis by the corresponding of P and Tr. This is not easy task in the spreadsheet so an extra table is built using the equations

$$P_L = e^{-e^{-y}} \text{ and } T = \frac{1}{P_R} = \frac{1}{1-P_L}$$

Exercise
For this exercise the maximum daily rainfall per year of RiskCity area is selected (1951-2001) N=50 yrs, see spreadsheet ExerciseFA-RiskCity.

1. Rank the yearly maximum rainfall values from low to high. So assign lowest rank 1 to the lowest data value and assign the highest rank N to the highest data value.
2. Calculate for each observation the left sided probabilities using Eq.1.
3. Determine the return period for each observation using Eq.2.
4. Determine the plotting position for each observation using Eq.3.
5. Fill in the box:

Rain for Tr = 10 years		Tr for rain = 65 mm	
Rain for Tr = 40 years		Tr for rain = 80 mm	
Rain for Tr = 100 years		Tr for rain = 90 mm	

6. Is there any outlier in the series?
7. Assuming that there is no error in the measurement, what does this "outlier" mean?
8. How can the outlier be treated?

Intensity-duration-frequency relationships

Example
Determine the design precipitation intensity and depth for a 20- minute duration storm with a 5-year return period for Figure 1.

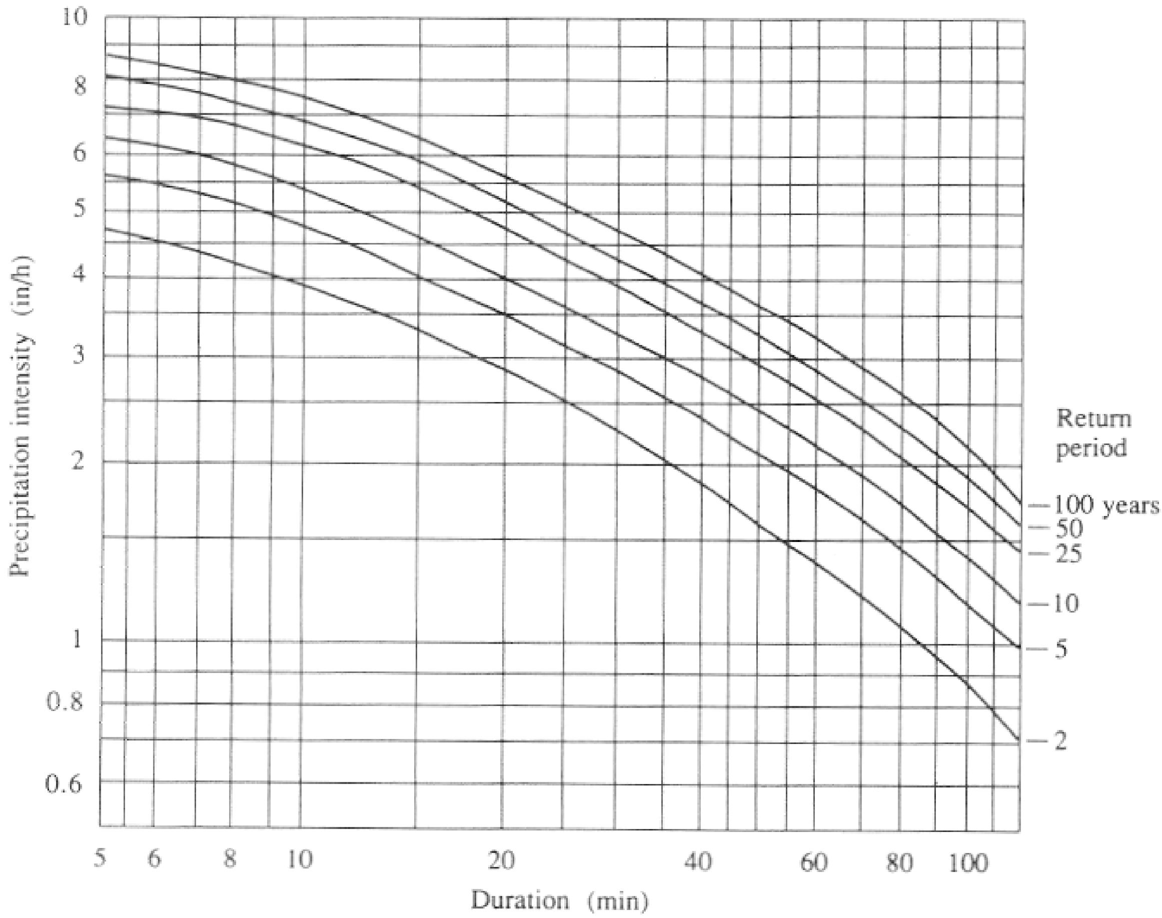


Figure 1: IDF curves are calculated for a certain station and it cannot be extrapolated to other areas.

Solution: From the IDF curves, the design intensity for a 5-year, 20-minute storm is i = 3.50 in/h. The corresponding precipitation depth is given by P= i*Td with Td = 20 min =0.333 h.
P= i*Td= 3.50* 0.333= 1.17 in.

Exercise
The data of RiskCity can be found in the spreadsheet: ExerciseFA-RiskCity.

1. Determine the design precipitation intensity and depth for 80- minute duration storm with a 20-year return period for the RiskCity area.
2. Determine the design precipitation intensity and depth for 10- minute duration storm with a 50-year return period for the RiskCity area.

Earthquakes

THE GUTENBERG-RICHTER RELATION

Concept: By plotting, on a logarithmic scale, the number of earthquakes greater than or equal to a given magnitude in a set period of time against that magnitude, a basic characteristic of the seismicity rate in an area -- the *b* value -- can be determined.

Procedure:
Beno Gutenberg and Charles Richter were two of the pioneers of modern seismology; each contributed greatly to the development of the field as a modern, quantitative science. In the 1930s, as instrumental recording of earthquakes was becoming a reality in many areas of the world, these two scientists described a pattern in the seismic data that related the number of earthquakes in a given area (or around the entire world) over a fixed period of time to the magnitude of those earthquakes. Using Richter's recently developed magnitude scale and the newest instrumental records, they found that the number of earthquakes greater than magnitude 6 that would occur in a given area over, say, 10 years, was proportional to the number of earthquakes greater than magnitude 5 in that area, which was proportional to the number greater than magnitude 4, and so on.

This activity consists of two exercises designed to familiarize you with the "Gutenberg-Richter relation", as the pattern described by these early seismologists came to be known. The exercises are outlined below. Each has its own set of instructions and review questions. Work through each as directed within the exercise itself.

In the first exercise you will be given a set of data to graph. Once you have determined which kind of graph to use and have plotted the data, it will be up to you to figure out the equation that describes the Gutenberg-Richter relation.

In the second exercise, you'll use the NEIC USGS earthquake catalog data to work on your own data set from Honduras.

Exercise 1 ◆ California and The World
Your introduction to Gutenberg-Richter plots will be a relatively easy one. The data will be provided; you only need to determine what sort of graph to make, and then plot the pre-made data sets. You'll start with a set of data from southern California, then plot worldwide totals of earthquakes against this and compare.

For southern California, the data set below was compiled according to the following guidelines:

- Start date: January 1, 1987; End date: December 31, 1996.
- "Southern California" is defined as that area between 32° and 36.25° North latitude, and 114.75° and 121° West longitude.
- Only earthquakes of magnitude 2.5 and greater were used, because the Southern California Seismic Network's (SCSN) database is almost certainly incomplete below this magnitude for the area, as outlined above.

Earthquake Numbers in Southern California, 1987 through 1996

Magnitude (M) Range	*Count per **M** Range*	*Cumulative Total Above Lower **M** in Range*
2.5 - 2.9	9471	13590
3.0 - 3.4	2784	4119
3.5 - 3.9	912	1335
4.0 - 4.4	285	423
4.5 - 4.9	90	138
5.0 - 5.4	32	48
5.5 - 5.9	10	16
6.0 - 6.4	3	6
6.5 - 6.9	2	3
7.0 - 7.4	1	1

Before you can do any graphing, you'll need to decide what type of graphing scale to use. Choose a simple x-y plot, with magnitude **M** as the x-axis and number of earthquakes greater than magnitude **M** as the y-axis.

Note that the x-axis data, the magnitudes, are very much linear in scale, increasing in half-unit steps. However, look how greatly our y-axis numbers change -- we'll need to plot the number 1 and the number 13590 on the same graph! If we used a proportional linear scale for each axis, the y-axis would be huge, while the x-axis would be miniscule!

But note that numbers we want to plot on the y-axis jump about a factor of ten for every unit in magnitude increase. This suggests that we could use a y-axis based on powers of 10, or a **logarithmic** scale, while we use a linear scale for the x-axis.

Introduce the data above in an excel sheet. You are now ready to begin making your first graph of the data set above. Do so now. Remember to pay attention to the scale for each axis, but don't worry too much about making your points exact. When you finish plotting the data, work through the questions below.

1. In roughly what form did the points you plotted fall? (What shape or pattern?) Or are they completely random, with no form at all?
2. You should see a roughly linear arrangement of points. Using a straight edge, draw a single line that best represents the set of points you've plotted. That line does not need to run through the centre, or even touch, all of the points in your set.

You now have a line that represents the data you graphed -- the numbers of earthquakes with respect to magnitude over 10 years in southern California (1987-1996). You are ready to describe the Gutenberg-Richter relation just as the two of them did, decades ago.

The equation for a line on a simple x-y plot is **y = bx + a**, where **a** is the y-intercept and **b** represents the slope of the line. To keep **b** positive at all times, we can think of the above equation as true for positive-sloping lines (going up as you move left-to-right), and the equation **y = a - bx** as true for negative-sloping lines.

3. Is the slope of your graphed line positive or negative?
4. Your graph has a y-axis that is logarithmic. Thus, a negative-sloping line on this graph would be described as **log y = a - bx**. Indeed, instead of calling it the y-axis, think of it as a function **N** of the magnitude, **M**, which itself can be substituted for x (since the x-axis is magnitude). Make these substitutions in the equation given above. What do you get?

You now have a mathematical expression that represents the Gutenberg-Richter relation, the correlation between the magnitude of earthquakes and their relative numbers. It should look something like

$$\log N(\mathbf{M}) = a - b\mathbf{M}$$

Had you come up with this 70 years earlier, this expression might have been named after you!

But do all data sets of earthquakes counted according to magnitude plot in this same linear manner? And even if they are all linear, does the slope of different sets vary significantly?

To begin to answer these questions, let's plot another set of data -- this time, the average values of an entire year's worth of worldwide seismicity. Using that same piece of graph paper you used to plot the southern California data set, plot the set of data given in the table below, then answer the questions that follow.

Average Worldwide Seismicity Totals for a Single Year

Magnitude (M)	*# Greater Than* ***M***
3.0	100000 +
4.0	15000
5.0	3000
6.0	100
7.0	20
8.0	2

5. Did this set of data plot as a line, too? If you haven't already, draw a line that is the best fit for these new data points.
6. How does the slope of this new line (worldwide seismicity) compare with that of the southern California data?

Use a ruler to actually measure the slope of each line you graphed. Pick any segment of each line and sketch out a right triangle, with the legs parallel to the axes, and the line itself forming the hypotenuse. Measure the height of the vertical side of the triangle and divide this by the length of the horizontal side. Your answer will be the slope of the line, otherwise known as the ***b* value**.

Another way to find the ***b* value** is to note the value of N(**M**) at each of two points along the line, exactly one magnitude unit (i.e. in the x-direction) apart. Divide the larger number (the point on the left) by the smaller number (the point on the right), and then take the logarithm of this quotient. That answer is the ***b* value** of this line.

7. Using whichever method you prefer, what numerical values do you get for the *b* value of each line?

As it turns out, when Gutenberg-Richter plots are made for various data sets all over the world, most end up having a *b* value very close to 1, usually slightly less. This basic relation seems to be a universal property of seismicity.

Exercise 2 ✦ Seismic hazard in Honduras
In this exercise, you will again be making a Gutenberg-Richter plot for Honduras. This time, however, you will not be provided a pre-formed data set; you must sort and bin the data yourself! The data is already collected for you (see Honduras_EQ.xls). You will need to use the histogram function (“tools > data analysis”). If not installed, please go to “tools > add-ins > and select “analysis toolpack”. Now you will find under tools the ‘data analysis’ option.

Do the following:
Create on your excel sheet a separate list of bin-ranges that you would like to use. You can use as example the bin-ranges as used in exercise 1.

- Select the histogram option from “tools > data analysis > histogram”.
- Select as input the data you would like to investigate (magnitude), select as bin-range the bin-values you have created, as output you can select a separate sheet or in your present sheet.
- Calculate Gutenberg-Richter relationship for the whole catalogue. If you haven't already, draw a line to best fit your set of graph points. Then measure the slope of this line. What is the *b* value of this data set (to the nearest tenth)? Compare this with the values observed for the World and California.
- Calculate now the Gutenberg-Richter relationship for:

    1. period till 1980
    2. period between 1980 till 1990
    3. period between 1990 and 2000
    4. period from 2000 till present

  Do you observe any differences for the various periods (in terms of completeness, overall seismic activity, *b* values?