Analysis Techniques: Flood Analysis Tutorial with Instaneous Peak Flow Data (Log-Pearson Type III Distribution)

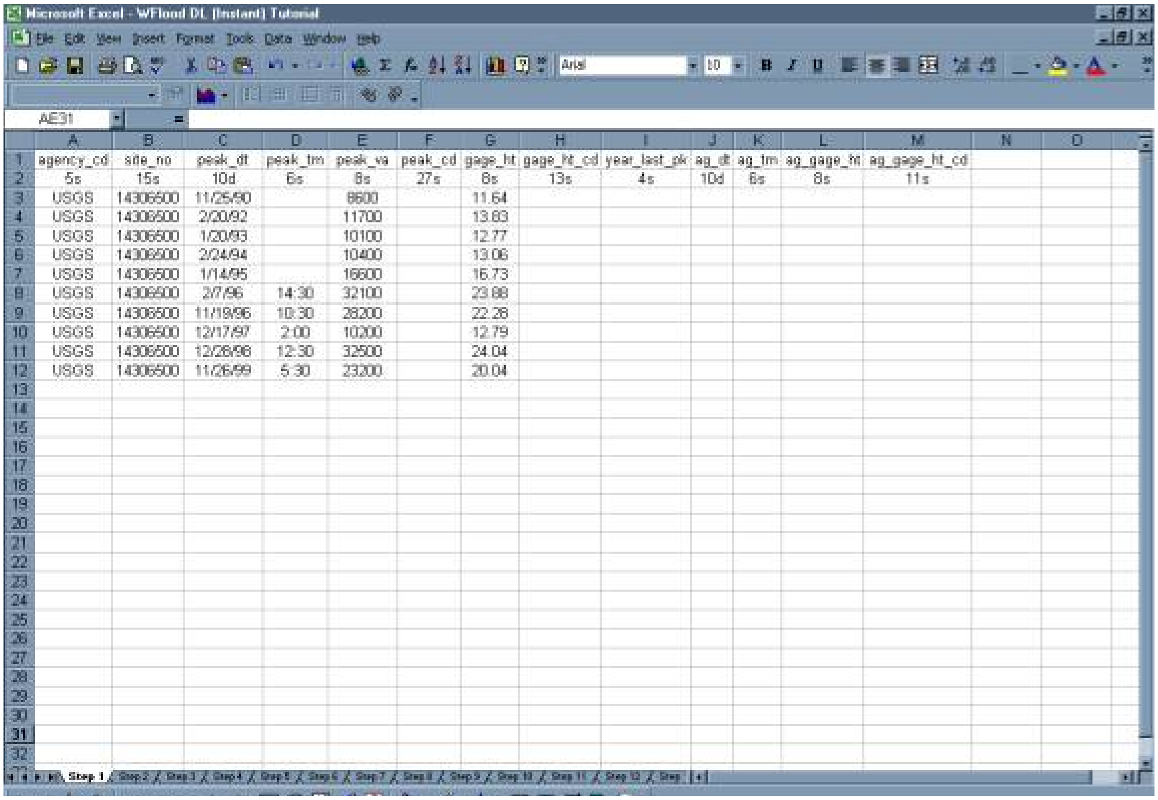
Information to get started:

- The lesson below contains step-by-step instructions and "snapshots" of what each step looks like when carried out in a Microsoft Excel workbook. Blue shading of information in the Excel illustrations denotes changes made from the previous step. Dots placed in three consecutive rows indicate that a portion of data is hidden from sight.
- You can download an Excel workbook containing the complete data set by clicking on the "Download Data" link below. It contains each calculation step on a separate worksheet. To move between steps, click on the tabs at the bottom of the excel window.
- When you download the file, it may open in your browser window. You may wish to use the "save as" function to save the file to a local drive and then reopen it in Excel. This will make it easier to flip between the online lesson and the example workbook.
- Finally, we want to remind you that the techniques explained on this site are statistically based; therefore results must be viewed as predictions and not as facts. Please use the techniques and the information obtained from them responsibly!

Download Data

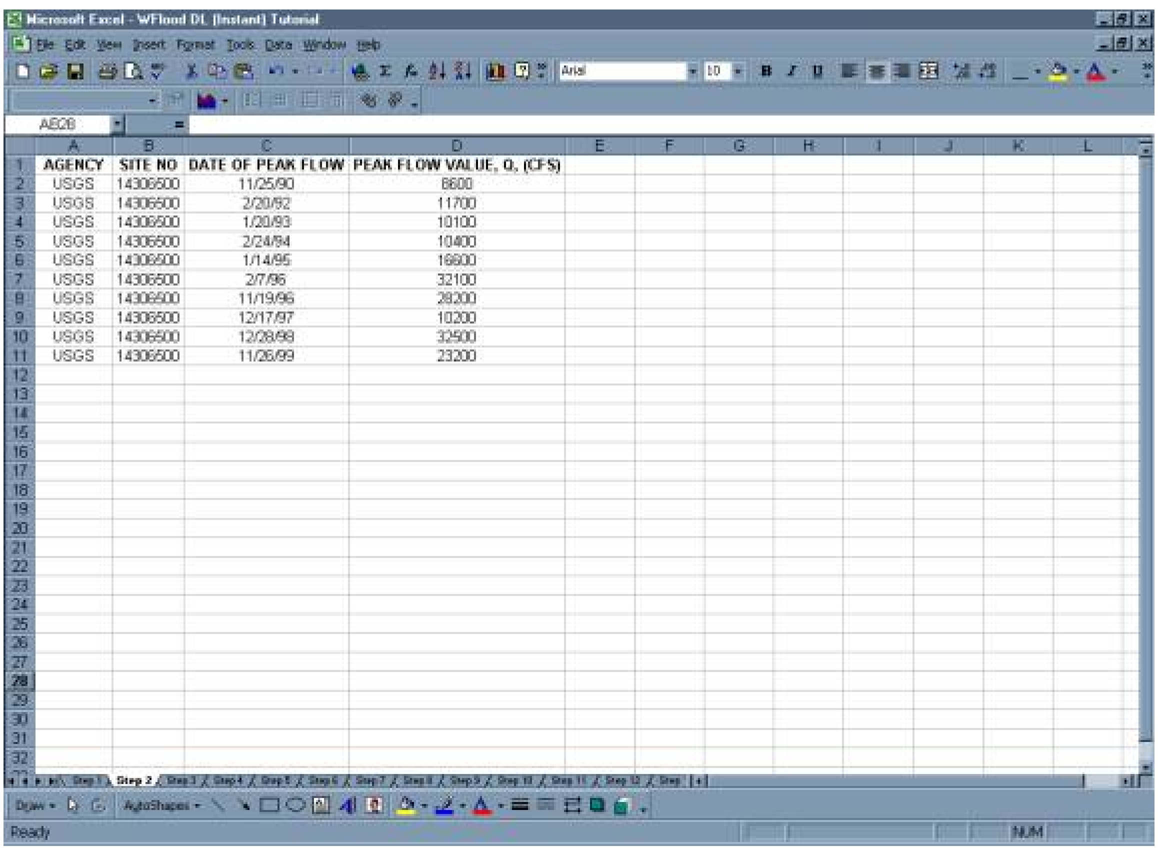
Step 1: Obtain streamflow data

- Obtain instantaneous peak streamflow data from the USGS web site.
  - Go to **http://oregon.usgs.gov**
  - Select Historical Water Data
  - Select Surface Water
  - Select Peaks
  - Check box under Site Identifier for Site Name and Submit
  - Type in Alsea under Site Name and select match any part and Submit
  - Select gage at TIDEWATER (140306500)
  - Select Tab-separated data
  - For the tutorial, copy the data for water years 1990 through 2000 into an Excel worksheet
  - Paste special as text (this will separate the data into columns)



agency_cd	site_no	peak_dt	peak_tm	peak_va	peak_cd	gage_ht	gage_ht_cd	year_last_pk	ag_dt	ag_tm	ag_gage_ht	ag_gage_ht_cd
5s	15s	10d	6s	8s	27s	8s	13s	4s	10d	6s	8s	11s
USGS	14306500	11/25/90		8600		11.64						
USGS	14306500	2/20/92		11700		13.83						
USGS	14306500	1/20/93		10100		12.77						
USGS	14306500	2/24/94		10400		13.06						
USGS	14306500	1/14/95		16600		16.73						
USGS	14306500	2/7/96	14:30	32100		23.88						
USGS	14306500	11/19/96	10:30	28200		22.28						
USGS	14306500	12/17/97	2:00	10200		12.79						
USGS	14306500	12/28/98	12:30	32500		24.04						
USGS	14306500	11/26/99	5:30	23200		20.04						

Step 2: Organize the information in a table.



AGENCY	SITE NO	DATE OF PEAK FLOW	PEAK FLOW VALUE, Q, (CFS)
USGS	14306500	11/25/90	8600
USGS	14306500	2/20/92	11700
USGS	14306500	1/20/93	10100
USGS	14306500	2/24/94	10400
USGS	14306500	1/14/95	16600
USGS	14306500	2/7/96	32100
USGS	14306500	11/19/96	28200
USGS	14306500	12/17/97	10200
USGS	14306500	12/28/98	32500
USGS	14306500	11/26/99	23200

Step 3: Rank the data from largest discharge to smallest discharge using the "sort" command. Add a column for Rank and number each streamflow value from 1 to n (the total number of values in your dataset).

Date of Peak Flow	Peak Flow Value, Q, (cfs)			Rank	Date of Peak Flow	Ranked Peak Flow Values, Q, (cfs)
11/25/90	8600			1	11/26/99	23200
2/20/92	11700			2	12/28/98	32500
1/20/93	10100			3	12/17/97	10200
2/24/94	10400			4	11/19/96	28200
1/14/95	16600			5	2/7/96	32100
2/7/96	32100			6	1/14/95	16600
11/19/96	28200			7	2/24/94	10400
12/17/97	10200			8	1/20/93	10100
12/28/98	32500			9	2/20/92	11700
11/26/99	23200			10	11/25/90	8600

Step 4: Create a column with the log of each max or peak streamflow using the Excel formula {log (Q)} and copy command.

Rank	Date of Peak Flow	Ranked Peak Flow Values, Q, (cfs)	log Q (cfs)
1	11/26/99	23200	4.365
2	12/28/98	32500	4.512
3	12/17/97	10200	4.009
4	11/19/96	28200	4.450
5	2/7/96	32100	4.507
6	1/14/95	16600	4.220
7	2/24/94	10400	4.017
8	1/20/93	10100	4.004
9	2/20/92	11700	4.068
10	11/25/90	8600	3.934

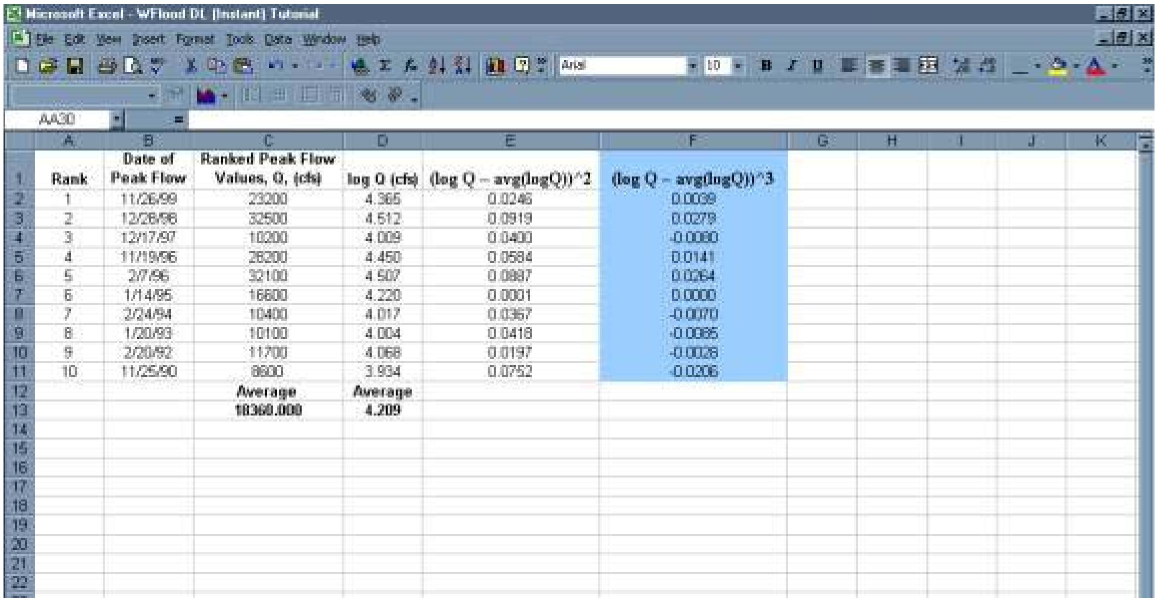
Step 5: Calculate the Average Max Q or Peak Q and the Average of the log (Q)

Rank	Date of Peak Flow	Ranked Peak Flow Values, Q, (cfs)	log Q (cfs)
1	11/26/99	23200	4.365
2	12/28/98	32500	4.512
3	12/17/97	10200	4.009
4	11/19/96	28200	4.450
5	2/7/96	32100	4.507
6	1/14/95	16600	4.220
7	2/24/94	10400	4.017
8	1/20/93	10100	4.004
9	2/20/92	11700	4.068
10	11/25/90	8600	3.934
		Average	**Average**
		18360.000	**4.209**

Step 6: Create a column with the excel formula {(log Q – avg(logQ))^2}

Rank	Date of Peak Flow	Ranked Peak Flow Values, Q, (cfs)	log Q (cfs)	(log Q – avg(logQ))^2
1	11/26/99	23200	4.365	0.0246
2	12/28/98	32500	4.512	0.0919
3	12/17/97	10200	4.009	0.0400
4	11/19/96	28200	4.450	0.0584
5	2/7/96	32100	4.507	0.0887
6	1/14/95	16600	4.220	0.0001
7	2/24/94	10400	4.017	0.0367
8	1/20/93	10100	4.004	0.0418
9	2/20/92	11700	4.068	0.0197
10	11/25/90	8600	3.934	0.0752
		Average	**Average**	
		18360.000	**4.209**	

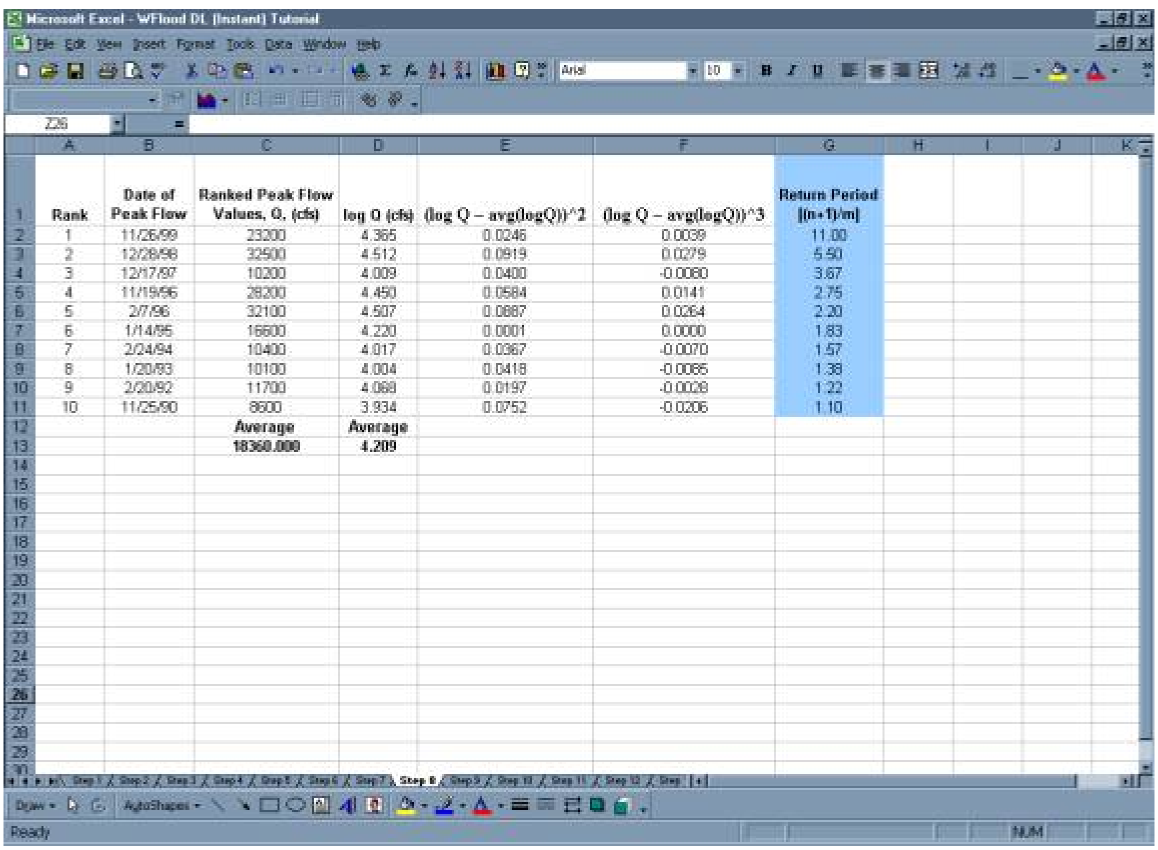
Step 7: Create a column with the excel formula {(log Q – avg(logQ))^3



Rank	Date of Peak Flow	Ranked Peak Flow Values, Q, (cfs)	log Q (cfs)	(log Q – avg(logQ))^2	(log Q – avg(logQ))^3
1	11/26/99	23200	4.365	0.0246	0.0039
2	12/28/98	32500	4.512	0.0919	0.0279
3	12/17/97	10200	4.009	0.0400	-0.0080
4	11/19/96	28200	4.450	0.0584	0.0141
5	2/7/96	32100	4.507	0.0887	0.0264
6	1/14/95	16600	4.220	0.0001	0.0000
7	2/24/94	10400	4.017	0.0367	-0.0070
8	1/20/93	10100	4.004	0.0418	-0.0085
9	2/20/92	11700	4.068	0.0197	-0.0028
10	11/25/90	8600	3.934	0.0752	-0.0206
		Average	Average		
		18360.000	4.209		

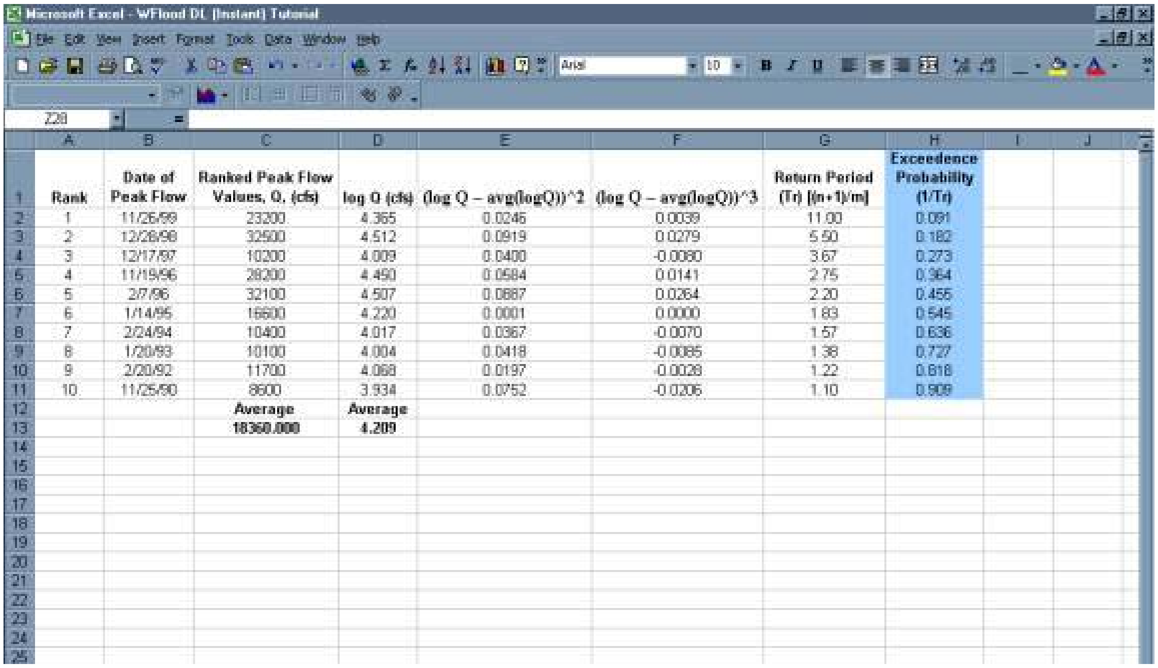
Step 8: Create a column with the return period (Tr) for each discharge using the Excel formula {(n+1)/m}.

Where n = the number of values in the dataset and m = the rank.



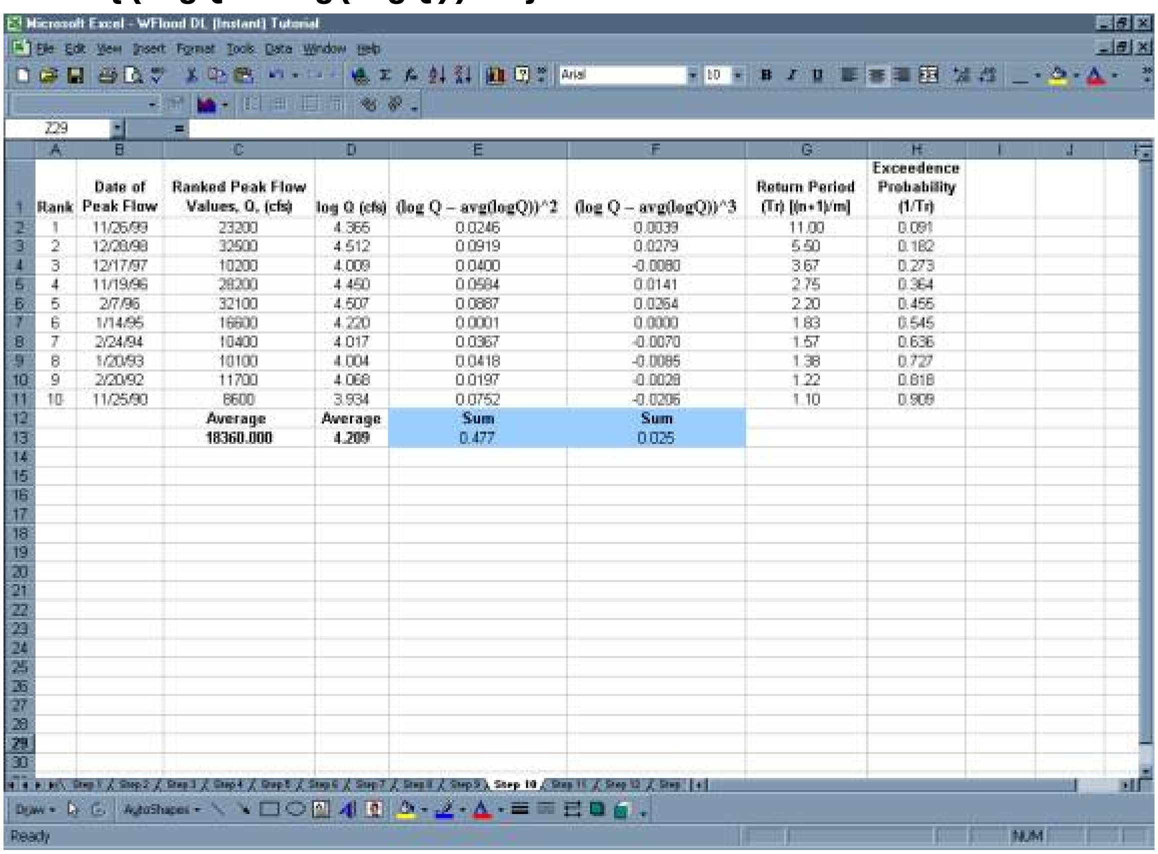
Rank	Date of Peak Flow	Ranked Peak Flow Values, Q, (cfs)	log Q (cfs)	(log Q – avg(logQ))^2	(log Q – avg(logQ))^3	Return Period [(n+1)/m]
1	11/26/99	23200	4.365	0.0246	0.0039	11.00
2	12/28/98	32500	4.512	0.0919	0.0279	5.50
3	12/17/97	10200	4.009	0.0400	-0.0080	3.67
4	11/19/96	28200	4.450	0.0584	0.0141	2.75
5	2/7/96	32100	4.507	0.0887	0.0264	2.20
6	1/14/95	16600	4.220	0.0001	0.0000	1.83
7	2/24/94	10400	4.017	0.0367	-0.0070	1.57
8	1/20/93	10100	4.004	0.0418	-0.0085	1.38
9	2/20/92	11700	4.068	0.0197	-0.0028	1.22
10	11/25/90	8600	3.934	0.0752	-0.0206	1.10
		Average	Average			
		18360.000	4.209			

Step 9: Complete the table with a final column showing the exceedence probability of each discharge using the excel formula {=1/Return Period or 1/Tr} and the copy command.



Rank	Date of Peak Flow	Ranked Peak Flow Values, Q, (cfs)	log Q (cfs)	(log Q – avg(logQ))^2	(log Q – avg(logQ))^3	Return Period (Tr) [(n+1)/m]	Exceedence Probability (1/Tr)
1	11/26/99	23200	4.365	0.0246	0.0039	11.00	0.091
2	12/28/98	32500	4.512	0.0919	0.0279	5.50	0.182
3	12/17/97	10200	4.009	0.0400	-0.0080	3.67	0.273
4	11/19/96	28200	4.450	0.0584	0.0141	2.75	0.364
5	2/7/96	32100	4.507	0.0887	0.0264	2.20	0.455
6	1/14/95	16600	4.220	0.0001	0.0000	1.83	0.545
7	2/24/94	10400	4.017	0.0367	-0.0070	1.57	0.636
8	1/20/93	10100	4.004	0.0418	-0.0085	1.38	0.727
9	2/20/92	11700	4.068	0.0197	-0.0028	1.22	0.818
10	11/25/90	8600	3.934	0.0752	-0.0206	1.10	0.909
		Average	Average				
		18360.000	4.209				

Step 10: Calculate the Sum for the {(logQ – avg(logQ))^2} and the {(logQ – avg(logQ))^3} columns.



Rank	Date of Peak Flow	Ranked Peak Flow Values, Q, (cfs)	log Q (cfs)	(log Q – avg(logQ))^2	(log Q – avg(logQ))^3	Return Period (Tr) [(n+1)/m]	Exceedence Probability (1/Tr)
1	11/26/99	23200	4.365	0.0246	0.0039	11.00	0.091
2	12/28/98	32500	4.512	0.0919	0.0279	5.50	0.182
3	12/17/97	10200	4.009	0.0400	-0.0080	3.67	0.273
4	11/19/96	28200	4.450	0.0584	0.0141	2.75	0.364
5	2/7/96	32100	4.507	0.0887	0.0264	2.20	0.455
6	1/14/95	16600	4.220	0.0001	0.0000	1.83	0.545
7	2/24/94	10400	4.017	0.0367	-0.0070	1.57	0.636
8	1/20/93	10100	4.004	0.0418	-0.0085	1.38	0.727
9	2/20/92	11700	4.068	0.0197	-0.0028	1.22	0.818
10	11/25/90	8600	3.934	0.0752	-0.0206	1.10	0.909
		Average	Average	Sum	Sum		
		18360.000	4.209	0.477	0.025		

Step 11: Calculate the variance , standard deviation , and skew coefficient as follows:

variance =

$$\frac{\sum_{i}^{n}(\log Q - avg(\log Q))^{\wedge}2}{n-1}$$

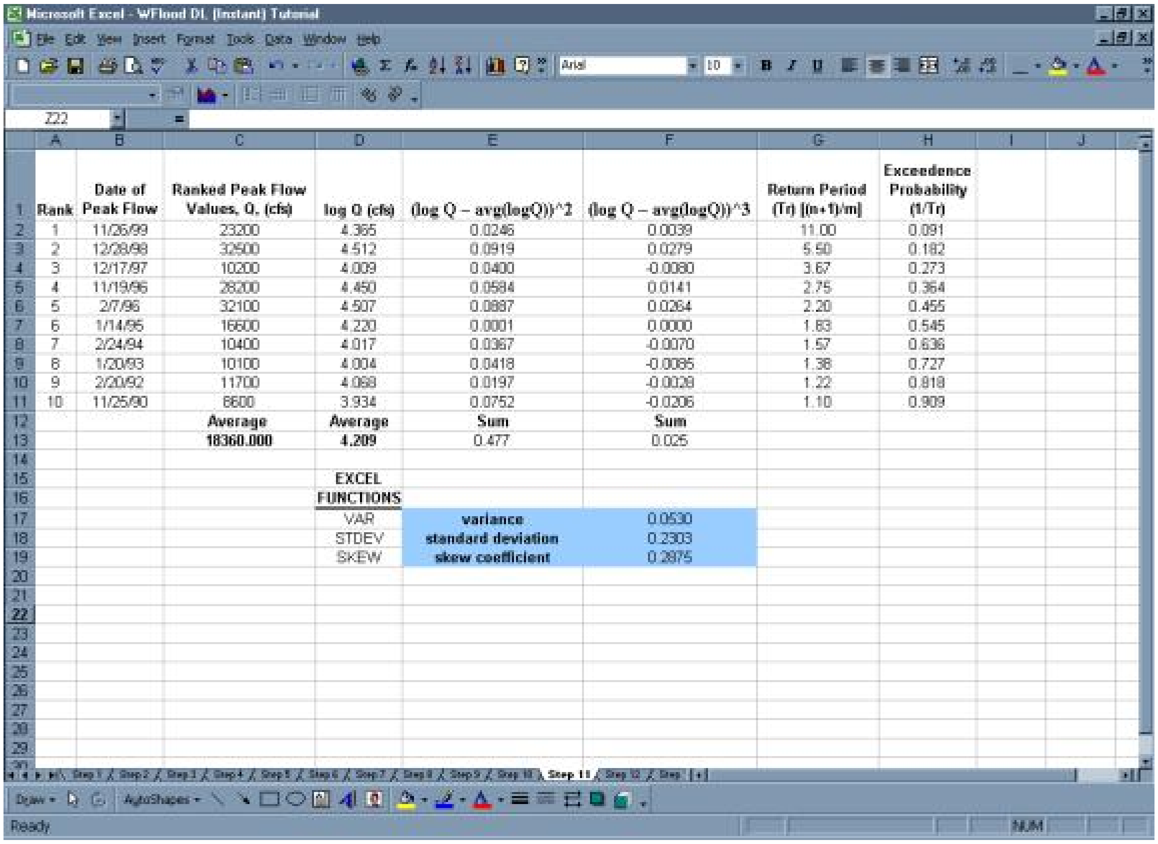
standard deviation =

$$\sigma \log Q = \sqrt{variance}$$

skew coefficient =

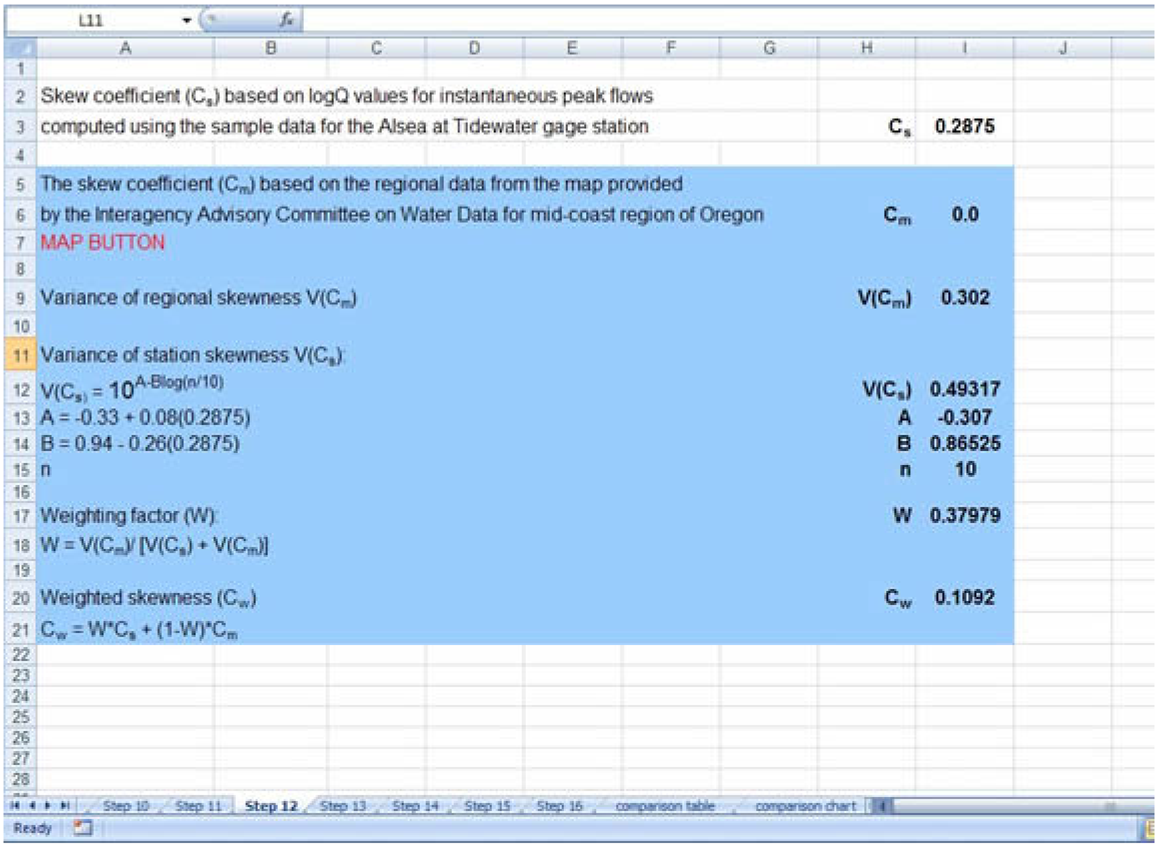
$$\frac{n \times \sum_{i}^{n}(\log Q - avg(\log Q))^{\wedge}3}{(n-1)(n-2)(\sigma \log Q)^{\wedge}3}$$

Excel functions can also be used to calculate the variance (=VAR( )), standard deviation (=STDEV( )), and skewness coefficient (=SKEW( )).



Rank	Date of Peak Flow	Ranked Peak Flow Values, Q, (cfs)	log Q (cfs)	(log Q – avg(logQ))^2	(log Q – avg(logQ))^3	Return Period (Tr) [(n+1)/m]	Exceedence Probability (1/Tr)
1	11/26/99	23200	4.365	0.0246	0.0039	11.00	0.091
2	12/28/98	32500	4.512	0.0919	0.0279	5.50	0.182
3	12/17/97	10200	4.009	0.0400	-0.0080	3.67	0.273
4	11/19/96	28200	4.450	0.0584	0.0141	2.75	0.364
5	2/7/96	32100	4.507	0.0887	0.0264	2.20	0.455
6	1/14/95	16600	4.220	0.0001	0.0000	1.83	0.545
7	2/24/94	10400	4.017	0.0367	-0.0070	1.57	0.636
8	1/20/93	10100	4.004	0.0418	-0.0085	1.38	0.727
9	2/20/92	11700	4.068	0.0197	-0.0028	1.22	0.818
10	11/25/90	8600	3.934	0.0752	-0.0206	1.10	0.909
		Average	Average	Sum	Sum		
		18360.000	4.209	0.477	0.025		
			EXCEL FUNCTIONS				
			VAR	variance	0.0530		
			STDEV	standard deviation	0.2303		
			SKEW	skew coefficient	0.2875		

Step 12: Calculate weighted skewness



	A	H	I
1			
2	Skew coefficient (C_s) based on logQ values for instantaneous peak flows		
3	computed using the sample data for the Alsea at Tidewater gage station	C_s	0.2875
4			
5	The skew coefficient (C_m) based on the regional data from the map provided		
6	by the Interagency Advisory Committee on Water Data for mid-coast region of Oregon	C_m	0.0
7	MAP BUTTON		
8			
9	Variance of regional skewness $V(C_m)$	$V(C_m)$	0.302
10			
11	Variance of station skewness $V(C_s)$:		
12	$V(C_s) = 10^{A-B\log(n/10)}$	$V(C_s)$	0.49317
13	$A = -0.33 + 0.08(0.2875)$	A	-0.307
14	$B = 0.94 - 0.26(0.2875)$	B	0.86525
15	n	n	10
16			
17	Weighting factor (W):	W	0.37979
18	$W = V(C_m) / [V(C_s) + V(C_m)]$		
19			
20	Weighted skewness (C_w)	C_w	0.1092
21	$C_w = W{*}C_s + (1-W){*}C_m$		

Show Me

Step 13: Calculate *K* values

- Use the **frequency factor table** and the skew coefficient to find the ***K* values** for the 2,5,10,25,50,100, and 200 **recurrence intervals**.
- If the skew coefficient is between two given skew coefficients in the table than you can linearly extrapolate between the two numbers to get the appropriate *K* value. To view the frequency factor table click on the "show me" link below.

Show Me

N22

	A	B	C	D	E	F	G	H	I
1	Rank	Date of Peak Flow	Ranked Peak Flow Values, Q, (cfs)	log Q (cfs)	(log Q – avg(logQ))^2	(log Q – avg(logQ))^3	Return Period (Tr) [(n+1)/m]	Exceedence Probability (1/Tr)	
2	1	11/26/1999	23200	4.365	0.0246	0.0039	11.00	0.091	
3	2	12/28/1998	32500	4.512	0.0919	0.0279	5.50	0.182	
4	3	12/17/1997	10200	4.009	0.0400	-0.0080	3.67	0.273	
5	4	11/19/1996	28200	4.450	0.0584	0.0141	2.75	0.364	
6	5	2/7/1996	32100	4.507	0.0887	0.0264	2.20	0.455	
7	6	1/14/1995	16600	4.220	0.0001	0.0000	1.83	0.545	
8	7	2/24/1994	10400	4.017	0.0367	-0.0070	1.57	0.636	
9	8	1/20/1993	10100	4.004	0.0418	-0.0085	1.38	0.727	
10	9	2/20/1992	11700	4.068	0.0197	-0.0028	1.22	0.818	
11	10	11/25/1990	8600	3.934	0.0752	-0.0206	1.10	0.909	
12			Average	Average	Sum	Sum			
13			18360.000	4.209	0.477	0.025			
14									
15				EXCEL					
16				FUNCTIONS					
17				VAR	variance	0.0530			
18				STDEV	standard deviation	0.2303			
19				SKEW	skew coefficient	0.2875			
20					weighted skew coefficient	0.1092			
21									
22				Tr	K(0.1)	K(0.2)	slope	K(0.1092)	Q (cfs)
23				2	-0.017	-0.033	-0.16	-0.018	
24				5	0.836	0.830	-0.06	0.835	
25				10	1.292	1.301	0.09	1.293	
26				25	1.785	1.818	0.33	1.788	
27				50	2.107	2.159	0.52	2.112	
28				100	2.400	2.472	0.72	2.407	
29				200	2.670	2.763	0.93	2.679	
30									
31									
32									

Step 10 | Step 11 | Step 12 | Step 13 | Step 14 | Step 15 | Step 16 | comparison table | comparison chart

Ready

Step 14: Using the general equation, list the discharges associated with each recurrence interval

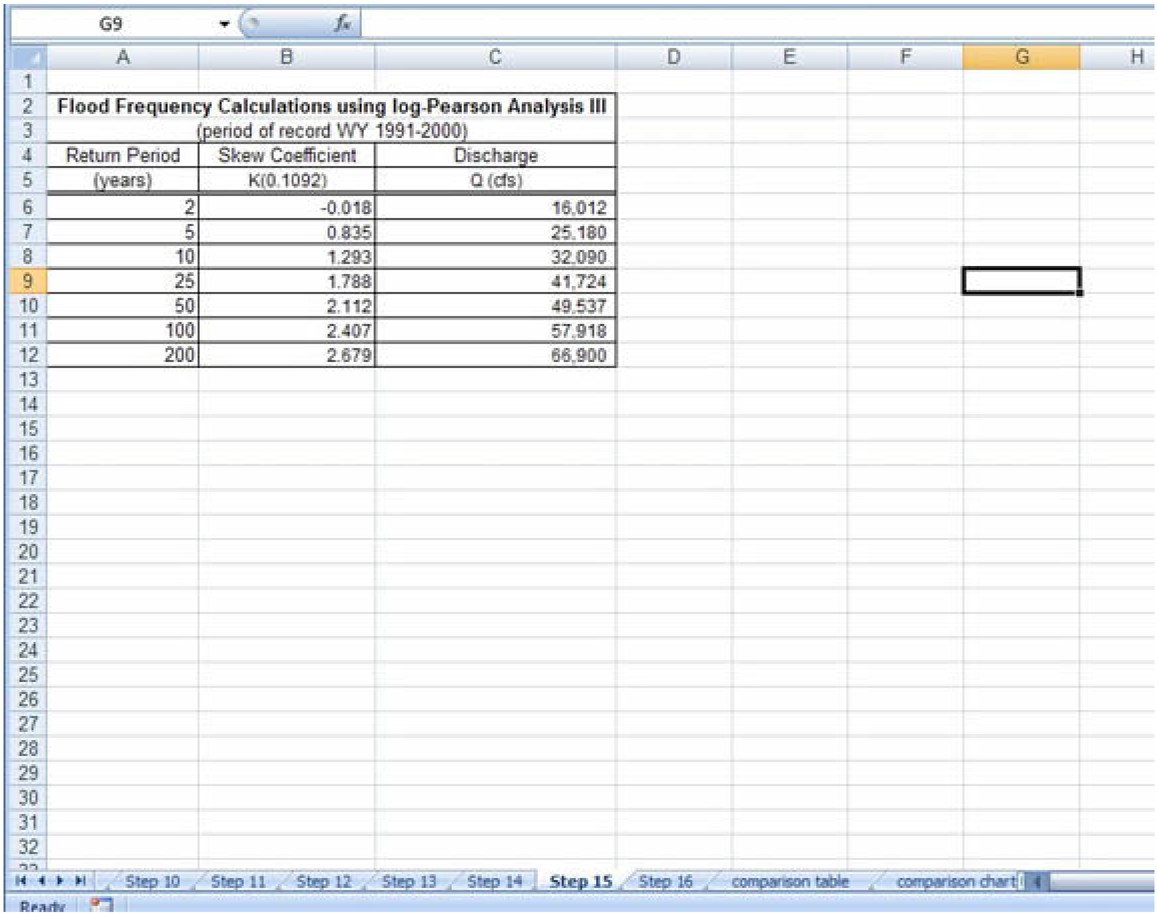
general equation =

$$\log QTr = avg(\log Q) + [K\,(Tr, Gs)] \times \sigma \log Q$$

	A	B	C	D	E	F	G	H	I
1	Rank	Date of Peak Flow	Ranked Peak Flow Values, Q, (cfs)	log Q (cfs)	(log Q – avg(logQ))^2	(log Q – avg(logQ))^3	Return Period (Tr) [(n+1)/m]	Exceedence Probability (1/Tr)	
2	1	11/26/1999	23200	4.365	0.0246	0.0039	11.00	0.091	
3	2	12/28/1998	32500	4.512	0.0919	0.0279	5.50	0.182	
4	3	12/17/1997	10200	4.009	0.0400	-0.0080	3.67	0.273	
5	4	11/19/1996	28200	4.450	0.0584	0.0141	2.75	0.364	
6	5	2/7/1996	32100	4.507	0.0887	0.0264	2.20	0.455	
7	6	1/14/1995	16600	4.220	0.0001	0.0000	1.83	0.545	
8	7	2/24/1994	10400	4.017	0.0367	-0.0070	1.57	0.636	
9	8	1/20/1993	10100	4.004	0.0418	-0.0085	1.38	0.727	
10	9	2/20/1992	11700	4.068	0.0197	-0.0028	1.22	0.818	
11	10	11/25/1990	8600	3.934	0.0752	-0.0206	1.10	0.909	
12			Average	Average	Sum	Sum			
13			18360.000	4.209	0.477	0.025			
14									
15				EXCEL					
16				FUNCTIONS					
17				VAR	variance	0.0530			
18				STDEV	standard deviation	0.2303			
19				SKEW	skew coefficient	0.2875			
20					weighted skew coefficient	0.1092			
21									
22				Tr	K(0.1)	K(0.2)	slope	K(0.1092)	Q (cfs)
23				2	-0.017	-0.033	-0.16	-0.018	16,012
24				5	0.836	0.830	-0.06	0.835	25,180
25				10	1.292	1.301	0.09	1.293	32,090
26				25	1.785	1.818	0.33	1.788	41,724
27				50	2.107	2.159	0.52	2.112	49,537
28				100	2.400	2.472	0.72	2.407	57,918
29				200	2.670	2.763	0.93	2.679	66,900

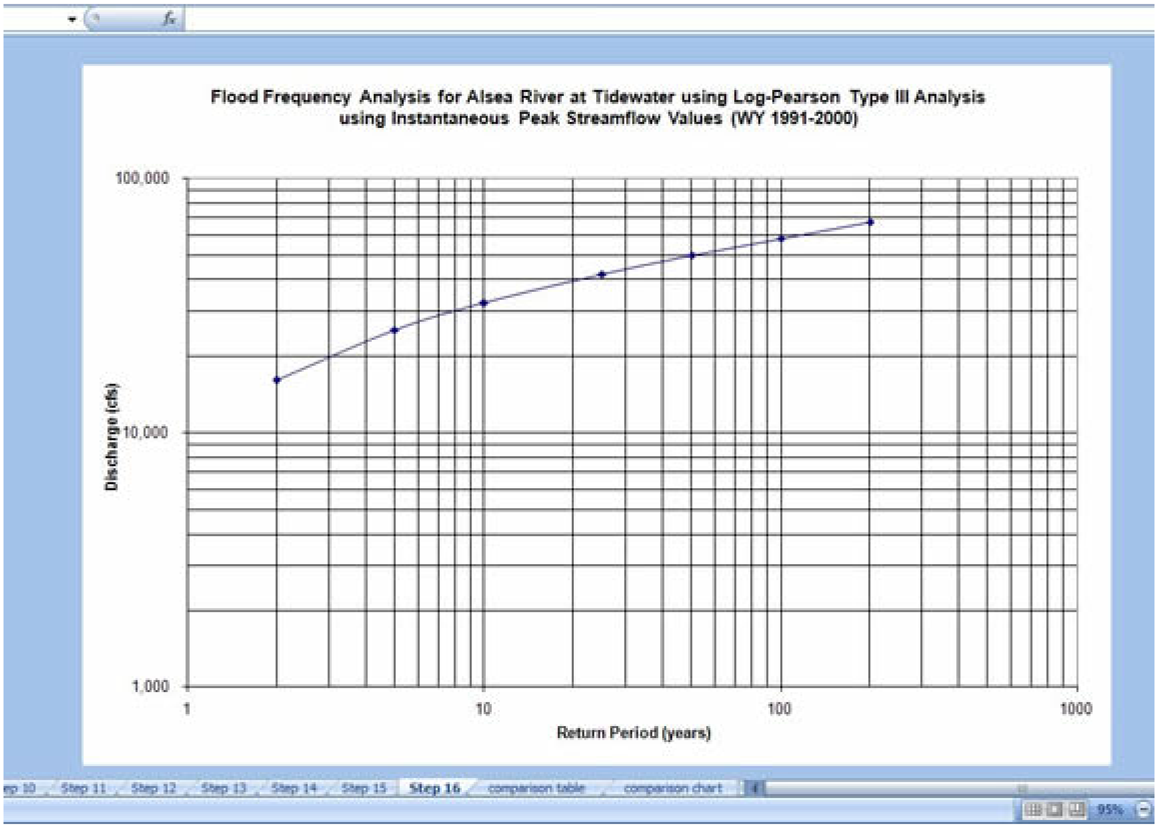
Step 10 | Step 11 | Step 12 | Step 13 | Step 14 | Step 15 | Step 16 | comparison table | comparison chart

Step 15: Create table of Discharge values found using the log – Pearson analysis



Flood Frequency Calculations using log-Pearson Analysis III (period of record WY 1991-2000)		
Return Period (years)	Skew Coefficient K(0.1092)	Discharge Q (cfs)
2	-0.018	16,012
5	0.835	25,180
10	1.293	32,090
25	1.788	41,724
50	2.112	49,537
100	2.407	57,918
200	2.679	66,900

Step 16: Create Plot



- Below is a comparison of flood frequency analysis completed using mean daily data versus instantaneous discharge data. As can be seen, had you completed this analysis using mean daily data, the result would have been an underestimation of the discharges, especially those associated with shorter return periods.