آزمون و بازسازی داده‌ها

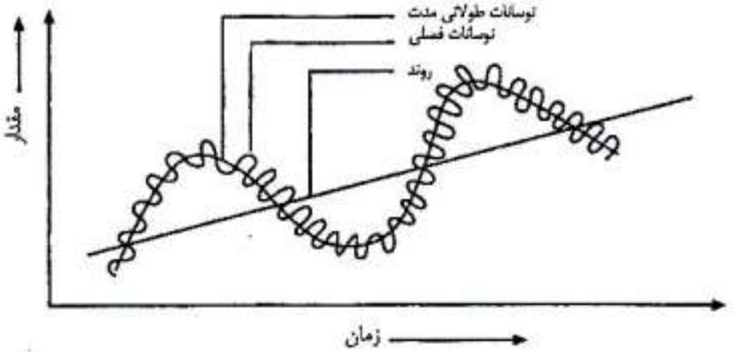
آزمون همگنی داده‌ها	**روش میانگین‌گیری**
روشهای نموداری	**روش نموداری**
روشهای غیر نموداری	**آزمون کفایت داده‌ها**
تخمین داده‌های غیر موجود	**آزمون روند داده‌ها**
روش درون‌یابی و برون‌یابی	**مسائل**
روش تفاضلها و نسبتها	**منابع برای مطالعه بیشتر**

۱۸-۱ آزمون همگنی داده‌ها

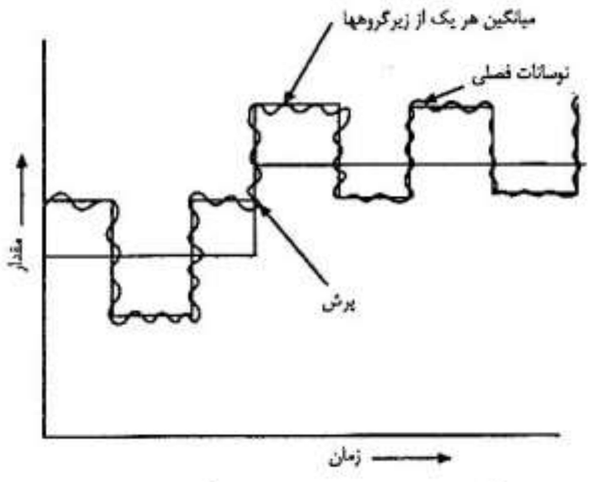
قبل از آن‌که متخصص هیدرولوژی به تجزیه و تحلیل داده‌ها اقدام کند لازم است از کیفیت آنها و همچنین کامل بودن سری آماری اطمینان حاصل نماید. بدون ارزیابی صحیح داده‌ها، انجام تحلیلهای پیچیدهٔ آماری نتایج قابل اعتمادی را بدست نخواهد داد. هرچند استاندارد بودن قرائت و ثبت داده‌های هیدرولوژی تا حد زیادی از ایجاد اشتباه در آنها می‌کاهد ولی معمولاً داده‌هایی که در اختیار هیدرولوژیست قرار می‌گیرد تا حدی به کنترل یا اصلاح نیاز دارند. عوامل مهمی که این اشتباهات را به وجود می‌آورند عبارتند از : تغییر محل ایستگاه‌های اندازه‌گیری، تغییر اشخاصی که کار قرائت را بر عهده دارند، منظم نبودن زمان اندازه‌گیری، تعویض ادوات ، احداث ساختمان در اطراف ایستگاه و غیره. تغییرات شدید موضعی در آب و هوای یک منطقه نیز چنین اشتباهاتی را به دنبال دارد و ممکن است باعث ناهمگنی داده‌ها شود.

اگر یک سری از داده‌های هیدرولوژیکی را در اختیار داشته باشیم و مقادیر آن‌ها را در یک دستگاه محور مختصات نسبت به زمان رسم کنیم ممکن است چند حالت را در آن‌ها مشاهده کنیم. یا آن که داده‌ها حول یک خط افقی قرار می‌گیرند که در این صورت گفته می‌شود سری هیدرولوژیکی ایستا (stationary) می‌باشد. در غیر این صورت ممکن است داده‌ها دارای نوسانات فصلی (seasonality)، نوسانات طولانی مدت (long oscillation) و یا روند (trend) باشند. تفاوت این تغییرات در شکل ۱۸-۱ نشان داده شده است. علاوه بر این ممکن است در خصوصیات حوضه

یا ایستگاهی که اندازه‌گیری شده است و یا شکل رودخانه تغییر ناگهانی صورت گرفته باشد. در این حالت به جای این که تغییر در داده‌های سری هیدرولوژیکی شکلی مشابه ۱۸-۱ داشته باشد پرش‌های ناگهانی (Jump) در آن‌ها مشاهده گردد (شکل ۱۸-۲). این اتفاقات ماهیت قطعی داشته و می‌توان مقادیر آن‌ها راکمی کرده و آن‌ها را برطرف کرد. بنابراین هیدرولوژیست باید از وجود یا عدم وجود چنین تغییراتی در داده‌ها آگاهی داشته باشد.



شکل ۱۸-۱ توسانات طولانی مدت، توسانات فصلی و روند در داده‌های هیدرولوژیکی



شکل ۱۸-۲ پیدایش پرش در میانگین داده‌ها

همگنی بمعنی این است که داده‌ها مربوط به یک جامعه آماری تصادفی مشخص باشند. برای این که ببینیم آیا داده‌ها و ارقامی که آن را تجزیه و تحلیل می‌کنیم همگن هستند یا خیر روشهای زیادی برای آزمون وجود دارد که می‌توان اصولاً آنها را در دو گروه دسته‌بندی کرد: یکی روشهای نموداری و دیگری روشهای غیر نموداری. با توجه به اهمیتی که این روشها در پردازش

داده‌های هیدرولوژیکی دارند به شرح مختصر آنها می‌پردازیم.

۱۸-۱-۱ روشهای نموداری

اگر یک سری از داده‌های هیدرولوژی را در اختیار داشته باشیم و هیچ‌گونه تغییرات دوره‌ای و یا غیر دوره‌ای در آنها وجود نداشته باشد به آنها داده‌های ثابت گفته می‌شود. در این صورت مقدار میانگین و واریانس در آنها ثابت است. به عبارت دیگر چنانچه داده‌های مشابه دیگری نیز به آنها اضافه شود نباید در میانگین و واریانس تغییری ایجاد نماید.

معمول‌ترین روش نموداری که به منظور اطمینان از یکنواختی داده‌ها بکار برده می‌شود روش به‌اصطلاح **جرم مضاعف** (double mass) است. طریقهٔ آزمون همگنی داده‌ها به روش جرم مضاعف برای یک ایستگاه فرضی (مثلاً A) به صورت زیر است:

(۱) داده‌های ایستگاه A را که قرار است مورد آزمون همگنی قرار گیرد به ترتیب سال وقوع ردیف نمایید.

(۲) چند ایستگاه (مثلاً B، C و D) را در اطراف ایستگاه A انتخاب کنید و میانگین داده‌های آنها را نیز مشابه ایستگاه A ردیف کنید. البته فرض می‌شود که آمار ایستگاههای B، C و D همگن و صحیح باشند.

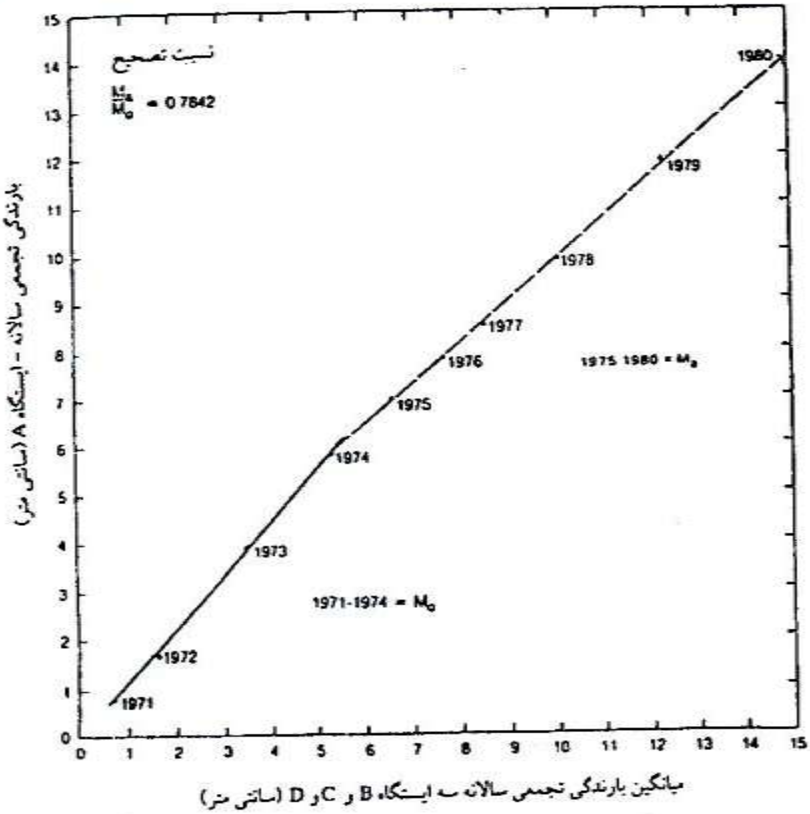
(۳) مقادیر تجمعی داده‌های ایستگاه A را برای هر سال محاسبه کنید.

(۴) مقادیر تجمعی میانگین ایستگاههای مجاور را نیز برای هرسال محاسبه نمایید.

(۵) در یک دستگاه محور مختصات، محور افقی را به داده‌های تجمعی‌میانگین ایستگاههای B، C و D و محور عمودی را به داده‌های تجمعی ایستگاه A اختصاص دهید و موقعیت مختصاتی هر یک از سالها را که بعنوان یک نقطه در نظر گرفته می‌شود مشخص کنید.

(۶) نقاط حاصله را بهم متصل کنید اگر از ابتدا تا انتها یک خط مستقیم حاصل شود در این صورت داده‌های ایستگاه A نسبت به ایستگاههای مجاور خود همگن است در غیر این صورت از هر زمان که تغییرات فاحشی در داده‌ها به وجود آمده باشد شیب خط تغییر کرده و دال بر عدم همگنی داده‌ها در طول دورهٔ آماری است.

در شکل ۱۸-۳ آمار بارندگی سالانه ایستگاه A مورد آزمون همگنی قرار گرفته است. سه ایستگاه B، C و D نیز در اطراف آن انتخاب شده و میانگین بارندگی سالانه آنها در همان سالهای آماری مشابه ایستگاه A محاسبه شده است. در شکل مذکور محور افقی بارندگی تجمعی سالانه میانگین ایستگاههای B، C و D و محور عمودی بارندگی تجمعی سالانه ایستگاه A را نشان می‌دهد که از سال ۱۹۷۱ الی ۱۹۸۰ موقعیت نقاط مربوط به آنها مشخص شده است. در این شکل مشاهده می‌شود که سالهای ۱۹۷۱ الی ۱۹۷۴ در امتداد یک خط و سالهای ۱۹۷۴



شکل ۱۸-۳ منحنی جرم مضاعف.

الی ۱۹۸۰ روی خط دیگری قرار می‌گیرند. بنابراین سالهای ۱۹۷۱ تا ۱۹۷۴ از یک روند و سالهای ۱۹۷۴ تا ۱۹۸۰ از روند دیگری پیروی می‌کنند. با فرض این که داده‌های مربوط به سالهای جدیدتر صحیح باشند باید آمار سالهای ۱۹۷۱ الی ۱۹۷۴ اصلاح شوند. چنانچه شیب خط مربوط به ۷۴-۱۹۷۱ برابر M_o و شیب خط ۸۰-۱۹۷۴ برابر M_a باشد، مقدار اصلاح شده داده‌ها از فرمول زیر به دست می‌آید.

$$p_a = \left(\frac{M_a}{M_o}\right)P_o \qquad (۱-۱۸)$$

که در آن:

P_o = داده‌های مشاهده شده در ایستگاه A

P_a = داده‌های اصلاح شده در ایستگاه A

M_a = شیب خطی که داده‌های مربوط به آن صحیح است (سالهای ۱۹۷۴ الی ۱۹۸۰)

M_o = شیب خطی که داده‌های آن باید اصلاح شود (سالهای ۱۹۷۱ الی ۱۹۷۴)

توجه شود که زمانی به اصلاح داده‌ها اقدام کنید که تغییر شیب بیش از ۱۰ درصد باشد.

● مثال ۱۸-۱

می‌خواهیم داده‌های ایستگاههای A ، B ، C ، D و E را در یک منطقه از نظر همگنی کنترل کنیم. این ایستگاهها از سال ۱۹۲۶ تا ۱۹۴۲ دارای آمار می‌باشند و مقدار سالانهٔ بارندگی هریک مطابق جدول ۱۸-۱ ردیف شده‌اند.

جدول ۱۸-۱

سال	A	B	C	D	E	میانگین
1926	39.75	45.70	3069	37.36	32.65	37.27
1927	39.57	38.52	40.99	30.87	28.08	35.61
1928	42.01	48.26	40.44	42.00	33.51	41.24
1929	41.39	34.61	32.49	39.92	29.58	35.60
1930	31.55	45.13	36.72	36.32	23.76	3470
1931	55.54	53.28	62.35	36.61	58.39	53.23
1932	48.11	40.08	47.85	38.61	46.24	44.18
1933	39.85	29.57	32.74	26.89	30.34	31.88
1934	45.40	41.68	36.13	32.44	46.78	40.49
1935	44.89	48.13	30.73	41.56	38.06	40.67
1936	32.64	35.48	35.40	31.32	42.82	36.33
1937	45.87	44.11	39.16	44.14	37.93	42.34
1938	46.05	38.94	43.27	50.62	50.67	45.91
1939	49.76	41.58	49.85	41.09	46.85	45.83
1940	47.26	49.66	47.86	39.01	50.52	16.86
1941	37.07	31.92	32.15	34.45	34.38	33.92
1942	45.89	38.16	52.39	47.32	47.60	46.27

حل

چون هیچکدام از ایستگاهها در مرحله اول مشکوک نمی‌باشند در جدول زیر میانگین بارندگی تجمعی هر کدام از ایستگاهها و نیز بارندگی تجمعی میانگین منطقه محاسبه شده است.

سال	A	B	C	D	E	میانگین
1926	39.75	45.70	30.69	37.36	32.85	37.27
1927	79.32	81.22	71.68	68.23	60.93	72.88
1928	121.33	132.48	112.12	110.23	94.44	114.12
1929	162.72	167.12	144.61	150.15	124.02	149.72
1930	194.27	212.25	181.33	186.47	147.78	184.42
1931	249.81	265.53	243.68	223.08	206.17	237.65
1932	297.92	305.61	291.53	261.69	252.41	281.83
1933	335.77	335.18	321.27	288.58	282.75	313.71
1934	383.17	370.86	360.40	321.02	329.53	354.20
1935	428.06	424.99	391.13	362.58	367.59	394.87
1936	460.70	464.47	426.53	303.98	410.41	431.20
1937	506.57	508.58	465.09	434.04	448.34	473.44
1938	552.62	547.52	508.96	488.66	499.01	919.35
1939	602.38	589.10	558.81	529.75	545.86	565.18
1940	649.61	638.76	606.67	568.76	596.38	612.04
1941	686.71	670.68	638.82	603.21	630.76	646.03
1942	732.60	708.81	691.21	650.53	678.36	692.30

مثلاً طرز محاسبه برای ایستگاه A به صورت زیر است.

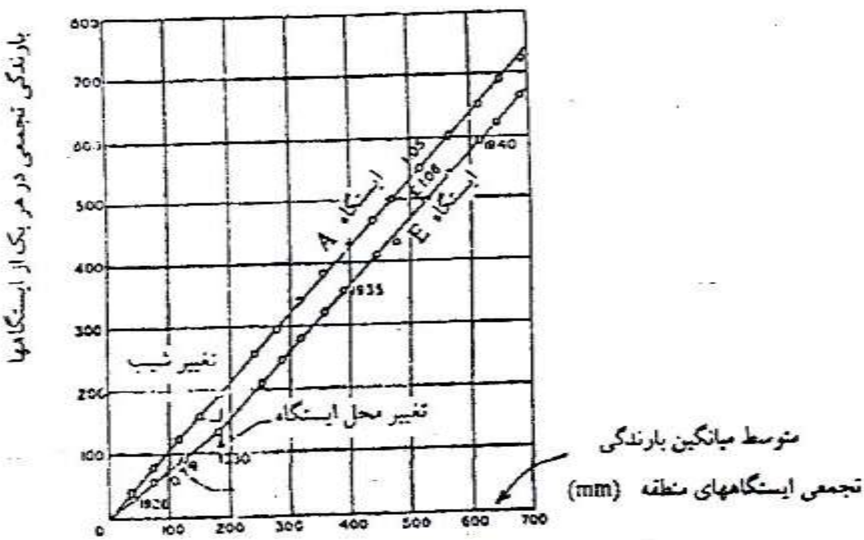
$39.75 =$ بارندگی در سال ۱۹۲۶

$39.75 + 39.57 = 79.32$ = بارندگی تجمعی تا سال ۱۹۲۷

$686.71 + 45.89 = 732.6$ = بارندگی تجمعی تا سال ۱۹۴۲

جمع میانگین بارندگی ایستگاههای منطقه (A، B، ... و E) منطقه نیز در ستون آخر سمت راست محاسبه و نوشته شده است. حال بارندگی تجمعی سالانه هریک از ایستگاهها را (ستون A، B، C، D و E) نسبت به میانگین بارندگی تجمعی تمام ایستگاهها (ستون آخر) در یک دستگاه محور مختصات رسم می‌کنیم. مثلاً اگر در محور عمودی بارندگی تجمعی ایستگاه A و در محور افقی میانگین بارندگی تجمعی منطقه آورده شود خطی مستقیم با شیب ۱/۰۵ مطابق شکل ۱۸-۲ به دست می‌آید. حاصل شدن خط مستقیم نشان دهندهٔ یکنواخت بودن آمار ایستگاه A است. اگر این عمل برای ایستگاههای B، C و D نیز انجام شود چنین خط مستقیمی حاصل می‌شود، لذا می‌توان گفت که داده‌های این ایستگاهها نیز یکنواخت است. ولی چنانچه این عمل برای ایستگاه E صورت گیرد مطابق شکل ۱۸-۴ مشاهده خواهد شد که از سال ۱۹۲۶ تا ۱۹۳۰ یک خط مستقیم با شیب ۰/۷۹ و از سال ۱۹۳۰ تا ۱۹۴۲ خط دیگری با شیب ۱/۰۶ حاصل می‌شود.

بدین ترتیب یکنواختی آمار ایستگاه E از سال ۱۹۳۰ تغییر کرده است و لذا یا ارقام سالهای ۱۹۲۶ تا ۱۹۳۰ باید تصحیح شوند و یا ارقام سالهای ۱۹۳۰ تا ۱۹۴۲. چون فرض براین است که آمار جدید صحیح‌تر از آمار قدیمی است، لذا ارقام مربوط به سالهای قدیمی‌تر یعنی ۱۹۲۶ تا ۱۹۳۰ تصحیح می‌گردند.



شکل ۱۸-۴ منحنی جرم مضاعف برای مثال ۱۸-۱

برای اصلاح کردن آمار از فرمول زیر استفاده می‌شود.

$$P_a = \frac{M_a}{M_o} P_o$$

در این مثال $M_a = ۱/۰۶$ ، $M_o = ۰/۷۹$ لذا ضریب تصحیح $\frac{M_a}{M_o} = \frac{۱/۰۶}{۰/۷۹} = ۱/۳۴$ است و برای تصحیح آمار باید داده‌های قبلی را در ۱/۳۴ ضرب کنیم.

بارندگی قدیم ایستگاه E در سال ۱۹۲۶ $P_o = ۳۲/۸۵$

بارندگی قدیم ایستگاه E در سال ۱۹۲۷ $P_o = ۲۸/۰۸$

بارندگی قدیم ایستگاه E در سال ۱۹۲۸ $P_o = ۳۳/۵۱$

بارندگی قدیم ایستگاه E در سال ۱۹۲۹ $P_o = ۲۹/۵۸$

بارندگی سال ۱۹۳۰ بدون تغییر باقی می‌ماند زیرا این نقطه روی هر دو خط است و لذا ضریب اصلاحی برای آن یک می‌باشد. مقادیر اصلاح شدهٔ بارندگیها عبارت خواهد بود از :

$$P_a(1926) = 32.85 \times 1.34 = 44.02$$

$$P_a(1927) = 28.08 \times 1.34 = 37.62$$

$$P_a(1928) = 33.51 \times 1.34 = 44.90$$

$$P_a(1929) = 29.58 \times 1.34 = 39.63$$

در مثال فوق ایستگاههای A ، B ، C ، D و E هرکدام بطور مجزا با میانگین کل که خود نیز شامل آن‌می‌شوند موردآزمون‌قرارگرفتند. ولی‌اگرآماریک ایستگاه خاص را بخواهیم بررسی کنیم می‌توان داده‌های‌مربوط‌به‌آن را با ایستگاههای شناخته شده و مطمئن دیگر مورد بررسی قرار داد.

● مثال ۱۸-۲

آماربارندگی‌سالانه‌ایستگاه A راازسال ۱۹۷۰ الی ۱۹۸۵ در اختیار داریم (جدول۱۸-۲) می‌خواهیم داده‌های این ایستگاه را بوسیله روش جرم مضاعف از نظر پایداری (consistancy) مورد بررسی قرار دهیم.

حل :

الف- داده‌های‌ایستگاه A بعنوان یک ایستگاه مشکوک درجدول ۱۸-۲ درستون ۲ ردیف شده است. ۴ ایستگاه B ، C ،D و E را در مجاورت این ایستگاه در نظر گرفته و آمار آنها را نیز مطابق جدول در ستون‌های ۳ تا ۶ ردیف می‌کنیم.

ب - میانگین بارندگی سالانه در ایستگاههای B ، C ،D و E را محاسبه می‌کنیم (ستون ۷).

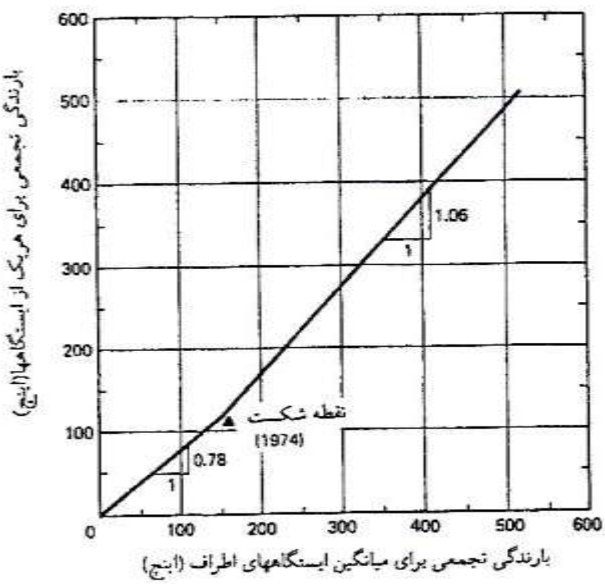
ج - بارندگی تجمعی ایستگاه A را برای هر سال محاسبه و در ستون ۸ قرار می‌دهیم.

د - بارندگی تجمعی میانگین‌ایستگاههای B ، C ،D و Eرامحاسبه‌ودر ستون ۹ قرار می‌دهیم.

ه - ارقام ستون ۸ (محور عمودی) را نسبت به ارقام ستون ۹ (محور افقی) در یک دستگاه محور مختصات مطابق شکل ۱۸-۵ رسم می‌کنیم.

جدول ۱۸-۲ بارندگی سالانه برای تحلیل جرم مضاعف

(1)	(2)	(3)	(4)	(5)	(6)	(7) میانگین ایستگاههای	(8) باران تجمعی در ایستگاه A	(9) باران تجمعی برای میانگین
	بارندگی سالانه در ایستگاه (in.)							
Year	A	B	C	D	E	B, C, D, E	Station A	B, C, D, E
1970	26.28	29.89	24.55	36.56	31.80	30.70	26.28	30.70
1971	22.46	24.70	32.79	30.82	31.66	29.99	48.74	60.69
1972	26.81	33.60	32.35	38.61	33.61	34.54	75.55	95.23
1973	23.66	31.94	25.99	27.71	33.11	29.69	99.21	124.92
1974	19.00	29.06	29.38	36.10	25.24	29.95	118.21	154.87
1975	46.71	29.29	49.88	42.62	44.43	41.56	164.92	196.43
1976	36.99	30.89	38.28	32.06	38.49	34.93	201.91	231.36
1977	24.27	21.51	26.19	23.66	31.88	25.81	226.18	257.17
1978	37.42	25.95	28.90	33.34	36.32	31.13	263.60	288.30
1979	30.45	33.25	24.58	38.50	35.91	33.06	294.05	321.36
1980	34.26	25.06	28.32	31.58	26.11	27.77	328.31	349.13
1981	30.34	35.31	31.33	35.29	36.70	34.66	358.65	383.79
1982	40.53	40.50	34.62	31.15	36.84	35.78	399.18	419.57
1983	37.48	32.87	39.88	33.26	39.81	36.46	436.66	456.03
1984	40.42	31.21	38.29	39.73	37.81	36.76	477.08	492.79
1985	27.50	27.56	25.72	25.54	29.66	27.12	504.58	519.91



شکل ۱۸-۵ منحنی جرم مضاعف برای مثال ۱۸-۲

و - در شکل مشاهده می‌شود که شیب خط رسم شده در ابتدا ۰/۷۸ و سپس در نقطه‌ای که مربوط به سال ۱۹۷۴ است شکسته شده وبه ۱/۰۶ میرسد لذا داده‌های سال‌های ۱۹۷۰ تا ۱۹۷۳ باید به نسبت $۱/۳۵۹ = \frac{۱/۰۶}{۰/۷۸}$ اصلاح شوند لذا بارندگی سالانه آنها را در ۱/۳۵۹ ضرب می‌کنیم تا اعداد زیر حاصل شود.

بارندگی اصلاح شده	ضریب اصلاح	بارندگی قبلی	سال
35.71	1.359	26.28	1970
30.52	1.359	22.46	1971
36.43	1.359	26.81	1972
32.15	1.359	23.66	1973

۱۸-۱-۲ روشهای غیر نموداری

روشهای گرافیکی از این جهت که معیار کمّی برای بیان حالت همگنی یا غیرهمگنی در آنها وجود ندارد روشهای کاملی به شمار نمی‌روند. روش ساده غیرنموداری که برای این منظور به کار برده می‌شود آزمون همگنی یا **ران تست** (run test) می‌باشد. در این روش به ترتیب زیر عمل می‌شود.

(۱) داده‌ها برطبق سال وقوع ردیف می‌شوند.

(۲) میانه یا میانگین داده‌ها به روشی که در فصل قبل گفته شد محاسبه و یکی از آنها را انتخاب می‌کنیم. فرض کنید میانگین را انتخاب کرده باشیم.

(۳) از ابتدای لیست داده‌ها که برحسب سال وقوعشان ردیف شده‌اند شروع می‌کنیم و هریک از داده‌ها که از نظر مقدار بالاتر از میانگین باشد با علامت a و هرکدام که پایین‌تر از میانگین باشد با b مشخص می‌نماییم (به خود عدد میانگین اگر در میان داده‌ها وجود داشته باشد علامتی تعلق نمی‌گیرد). مسلم است که اگر از میانه استفاده شده باشد تعداد aها و b هابرابر خواهند بود ($n_a = n_b$) ولی در صورتی که میانگین به عنوان عدد وسط در نظر گرفته شود ممکن است n_a (تعداد a ها) و n_b (تعداد b ها) برابر نباشد.

(۴) به ترتیب تعداد a ها و b ها را شمارش می‌کنیم تاn_a و n_b بدست آید. همچنین تعداد دنباله‌های aو b را نیز شمارش کرده و مجموع هر دو نوع دنباله را با حرف U مشخص می‌کنیم. منظور از دنباله، یک a یا یک سری aهای پشت سر هم و یک b یا bهای پشت سر هم است. زیرا مسلما این حروف دقیقاً بصورت یک درمیان اتفاق نخواهند افتاد و ممکن است برای مواردی چندین a یا b پشت‌سرهم قرار گیرند.

(۵) از جدول ۱۸-۳ با توجه به تعداد n_a و n_b اگر عدد U بین ارقام نوشته شده در جدول قرار گیرد داده‌ها همگن می‌باشند. یعنی تصادفی بودن آنها در سطح اعتماد ۹۵ درصد مورد قبول است. در این جدول هریک از n_a یا n_b که بزرگتر بودند در ردیف افقی بالای جدول و هرکدام

کوچکتر بود در ستون عمودی سمت چپ آورده شده و دو خط عمودی و افقی از این نقاط اخراج می‌شود. در محل تلاقی این خطوط دو عدد نوشته شده است که شرط تصادفی بودن داده‌ها آن است که U بین این دو عدد قرار گیرد.

جدول ۱۸-۳ حدود مجاز U

na یا nb (عدد کوچکتر)	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
2								2 6	2 6	2 6	2 6	2 6	2 6	2 6	2 6	2 6
3		2 8	2 8	2 8	2 8	2 8	2 8	2 8	2 8	2 8	3 8	3 8	3 8	3 8	3 8	3 8
4	2 9	2 9	2 10	3 10	3 10	3 10	3 10	3 10	3 10	3 10	3 10	4 10	4 10	4 10	4 10	4 10
5	2 10	3 10	3 11	3 11	3 12	3 12	4 12	4 12	4 12	4 12	4 12	4 12	4 12	5 12	5 12	5 12
6		3 11	3 12	3 12	4 13	4 13	4 13	4 13	5 14	5 14	5 14	5 14	5 14	5 14	6 14	6 14
7			3 13	4 13	4 14	5 14	5 14	5 14	5 15	5 15	6 15	6 16	6 16	6 16	6 16	6 16
8				4 14	5 14	5 15	5 15	6 16	6 16	6 16	6 16	6 17	7 17	7 17	7 17	7 17
9					5 15	5 16	6 16	6 16	6 17	7 17	7 18	7 18	7 18	8 18	8 18	8 18
10						6 16	6 17	7 17	7 18	7 18	7 18	8 19	8 19	8 19	8 20	9 20
11							7 17	7 18	7 19	8 19	8 19	8 20	9 20	9 20	9 21	9 21
12								7 19	8 19	8 20	8 20	9 21	9 21	9 21	10 22	10 22
13									8 20	9 20	9 21	9 21	10 22	10 22	10 23	10 23
14										9 21	9 22	10 22	10 23	10 23	11 23	11 24
15											10 22	10 23	11 23	11 24	11 24	12 25
16												11 23	11 24	11 25	12 25	13 25
17													11 25	12 25	12 26	13 26
18														12 26	13 26	13 27
19															13 27	13 27
20																14 28

مثال: به ازای $n_a = 12$ و $n_b = 12$ حدود مجاز U بین ۷ تا ۱۹ است.

● مثال ۱۸-۳

بارندگی فروردین‌ماه ایستگاه هواشناسی A از سال ۱۳۴۰ الی ۱۳۶۴ به مدت ۲۵ سال در اختیار می‌باشد. می‌خواهیم این داده‌ها را مورد آزمون قرار داده و ببینیم آیا داده‌های تصادفی هستند یا خیر؟ زیرا شرط پردازش به آنها حصول اطمینان از تصادفی بودن آنهاست. داده‌های مذکور در ستون دوم جدول ۱۸-۴ نوشته شده است.

حل

برای آزمون همگنی یک نقطه حد وسط را انتخاب می‌کنیم. برای این منظور میانه یا میانگین را در نظر می‌گیریم.

الف – در صورتی که میانه را در نظر بگیریم :

۱– چنانچه میانهٔ داده‌های ستون ۲ را حساب کنیم ۴۴ خواهد شد. حال به هر عدد که بالاتر از ۴۴ باشد علامت a و به هر عدد که کوچکتر از آن باشد علامت b داده می‌شود. به عدد ۴۴ علامتی اختصاص داده نشده است (جدول ۱۸–۴ ستون ۳).

۲– تعداد b و a ها شمارش می‌شوند. در این مثال،

$n_a = 12$ (تعداد aها)

$n_b = 12$ (تعداد bها)

$U = 12$ (جمع تعداد دنباله‌های a و b)

۳– در جدول ۱۸–۳ در ردیف افقی بالا عدد ۱۲ (یا هرکدام از n_a و n_b که بزرگتر باشند) و در ستون سمت چپ نیز عدد ۱۲ (یا هرکدام از n_a و n_b که کوچکتر باشد) را پیدا کرده در محل تلاقی خطوط عمودی و افقی که از این نقاط اخراج شود دو عدد مشاهده می‌شود که عبارتند از ۷ و ۱۹ حال چون U (یعنی ۱۲) بین این دو عدد قرار گرفته است، لذا در سطح اعتماد ۰/۰۵ این داده‌ها تصادفی می‌باشند.

ب – در صورتی که میانگین را در نظر بگیریم

۱– میانگین بارندگیها ۴۷/۸ میلی‌متر است که اگر داده‌ها را براساس این رقم علامت‌گذاری کنیم ستون ۴ حاصل می‌شود، a مربوط به ارقام بزرگتر از ۴۷/۸ و b مربوط به اعداد کوچکتر از آن است (جدول ۱۸–۴ ستون ۴).

۲– تعداد دنباله‌ها براساس این طبقه‌بندی به صورت زیر است :

$n_a = 11$ (تعداد aها)

$n_b = 14$ (تعداد bها)

$U = 12$ تعداد دنباله‌ها

۳– از جدول ۱۸–۳ استفاده می‌کنیم و n_b را در ردیف افقی بالای جدول و n_a را در ستون عمودی سمت چپ می‌آوریم. در محل تلاقی خطوط عمودی و افقی که از نقاط ۱۴ و ۱۱ رسم می‌شود دو عدد نوشته شده است که عبارتند از ۸ و ۱۹ و برای تصادفی بودن داده‌ها U باید در بین این دو عدد قرار گیرد که ملاحظه می‌شود وضعیت بهمین صورت است. مشاهده می‌شود که داده‌های بارندگی ایستگاه A به لحاظ آماری، همگن است. اصولاً آزمون همگنی قبل از تجزیه و تحلیل آماری روی داده‌ها، باید انجام شود تا از تصادفی بودن داده‌ها اطمینان حاصل گردد.

جدول ۱۸-۴

۱	۲	۳		۴	
سال	بارندگی (mm)	دنباله‌ها براساس میانه		دنباله‌ها بر اساس میانگین	
۱۳۴۰	۳۲	b	۱	b	۱
۱۳۴۱	۴۵	a	۲	b	
۱۳۴۲	۴۹	a		a	۲
۱۳۴۳	۴۲	b	۳	b	۳
۱۳۴۴	۳۰	b		b	
۱۳۴۵	۶۷	a	۴	a	۴
۱۳۴۶	۷۲	a		a	
۱۳۴۷	۷۳	a		a	
۱۳۴۸	۷۹	a		a	
۱۳۴۹	۲۵	b	۵	b	۵
۱۳۵۰	۲۸	b		b	
۱۳۵۱	۳۶	b		b	
۱۳۵۲	۴۹	a	۶	a	۶
۱۳۵۳	۵۲	a		a	
۱۳۵۴	۲۱	b	۷	b	۷
۱۳۵۵	۴۲	b		b	
۱۳۵۶	۴۹	a	۸	a	۸
۱۳۵۷	۶۷	a		a	
۱۳۵۸	۴۳	b	۹	b	۹
۱۳۵۹	۷۵	a	۱۰	a	۱۰
۱۳۶۰	۲۱	b	۱۱	b	۱۱
۱۳۶۱	۱۷	b		b	
۱۳۶۲	۴۴	–		b	
۱۳۶۳	۴۲	b		b	
۱۳۶۴	۹۶	a	۱۲	a	۱۲

۱۸-۲ تخمین داده‌های غیر موجود

یکی از کاربردهای آمار در هیدرولوژی آن است که بتوانیم برخی خصوصیات آب و هوایی یا هیدرولوژیکی مناطقی را که دارای داده‌های آماری کم یا اصولاً فاقد آمار هستند تخمین بزنیم. زیرا در بسیاری موارد نمی‌توان اجرای یک پروژه را فقط به دلیل این که در مورد آن داده‌های هیدرولوژیکی دراز مدت وجود ندارد به تعویق انداخت. از طرف دیگر نمی‌توان نقش داده‌ها رادر طراحیهای هیدرولیکی نادیده گرفت. بدین جهت لازم است به روشهای مختلف داده‌های مورد نیاز را تخمین زد. روشهای تخمین داده‌های هیدرولوژی در مورد تکمیل سریهای آماری نیز به‌کار برده می‌شود زیرا موارد زیادی مشاهده می‌شود که بنا به دلایلی آمار یک روز یا یک ماه و سال بخصوص مفقود شده یا اصولاً برداشت نشده است. این داده‌ها نیز قبل از آن‌که مورد تجزیه و تحلیل قرارگیرند باید تکمیل شوند. روشهایی که برای این منظور به‌کار می‌روند متعدد است. پاره‌ای از روشها گرافیکی و برخی روشها غیرگرافیکی می‌باشند که ما در این جا به تشریح برخی از آنها می‌پردازیم.

۱۸-۲-۱ روش درون‌یابی و برون‌یابی

ساده‌ترین روش برای تخمین داده‌های مفقود شده آن است که چند نقطه دیگر را که دارای آمار می‌باشند در اطراف نقطه مورد نظر انتخاب کنیم و با توجه به فاصله‌ای که دارند خطوط تراز آن منطقه را رسم نماییم و مشاهده کنیم که چه خطی از نقطهٔ مورد نظر می‌گذرد. مثلاً اگر باران سال ۱۳۶۰ ایستگاه A را نداشته باشیم از روی داده‌های ایستگاههای مجاور و با توجه به درون‌یابی بین داده‌های موجود می‌توان خطوط همباران سال ۱۳۶۰ را رسم و مشاهده کرد که کدام خط از نقطه A می‌گذرد. بارندگی نقطهٔ A برابر بارندگی خط همبارانی است که از آن نقطه عبور می‌کند. یکی از دلایل رسم خطوط همباران یا همدما نیز همین موضوع می‌باشد.

۱۸-۲-۲ روش تفاضلها و نسبتها

فرض کنید آمار بارندگی ایستگاه A از سال ۱۳۵۵ تا ۱۳۶۵ دراختیار است و می‌خواهیم آمار موجود نبوده آن را برای سالهای ۱۳۵۰ تا ۱۳۵۴ و مفقود شده آن را برای سال ۱۳۵۹ تخمین بزنیم (ستون ۲ جدول ۱۸-۵). یک ایستگاه را که آمار آن کامل و مورد اطمینان باشد در همان منطقه به عنوان ایستگاه مبنا انتخاب و فرض کنید، ایستگاه B به عنوان ایستگاه مبنا باشد که آمار آن در ستون ۳ نوشته شده است. مقایسهٔ این دو ایستگاه نشان می‌دهد که در سالهای ۱۳۵۵ تا ۱۳۵۸ و ۱۳۶۰ تا ۱۳۶۵ هر دو ایستگاه دارای آمار می‌باشند، چنانچه:

X_B = میانگین بارندگی در سالهای مشترک آماری (سالهایی که هر دو ایستگاه دارای آمار هستند) برای ایستگاه مبنا

X_A = میانگین بارندگی در سالهای مشترک آماری برای ایستگاه A باشد، مقدار بارندگی برای هریک از سالهای فاقد آمار در ایستگاه A به شرح زیر محاسبه می‌شود:

بارندگی در سال ۱۳۵۰ $X_1 = 310 \times \frac{X_A}{X_B}$

بارندگی درسال ۱۳۵۱ $X_2 = 352 \times \frac{X_A}{X_B}$

بارندگی در سال ۱۳۵۹ $X_6 = 300 \times \frac{X_A}{X_B}$

جدول ۱۸-۵

1	2	3
سال	بارندگی سالانه ایستگاه A(mm)	بارندگی سالانهٔ ایستگاه مبنا یا B(mm)
1350	-	310
1351	-	352
1352	-	381
1353	-	430
1354	-	460
1355	280	370
1356	256	282
1357	255	251
1358	230	293
1359	-	300
1360	382	310
1361	310	365
1362	280	393
1363	250	310
1364	262	280
1365	242	275

میانگین کل بارندگی در ایستگاه A نیز برابر است با:

میانگین کل در A = [میانگین سالهای مشترک آماری در A / میانگین سالهای مشترک آماری در B] × (میانگین کل در B)

مثلاً برای ایستگاه A در سال ۱۳۵۰ باید رقم ۲۷۲ را قرار داد زیرا،

میانگین بارندگی در سالهای مشترک در A = 274.7

میانگین بارندگی در سالهای مشترک در B = 312.9

بارندگی سال ۱۳۵۰ در A = $310 \times \frac{274.7}{312.9} = 272.1$

و یا چون میانگین کل بارندگی در ایستگاه مبنا (B) برای تمام سالها ۳۳۵/۱ میلی‌متر است، لذا میانگین کل بارندگی ایستگاه A عبارت خواهد بود از :

$$\text{میانگین کل بارندگی در ایستگاه A} = 335.1 \left(\frac{274.7}{312.9}\right) = 294.1$$

روش نسبتها بیشتر برای داده‌های بارندگی و رواناب به کار برده می‌شود. برای تکمیل داده‌های دما و مشابه آن از روش تفاضلها استفاده می‌شود که اساساً مشابه روش نسبتها می‌باشد. برای روشن شدن طرز استفاده از روش تفاضلها به ذکر یک مثال می‌پردازیم.

● مثال ۱۸-۴

آمار حداکثر دمای سالانه در ایستگاه A به مدت ۱۱ سال بطور کامل در اختیار است. ایستگاه B فقط به مدت ۷ سال دارای آمار می‌باشد. درجهٔ حرارت را برای ایستگاه B در سالهایی که فاقد آمار است تخمین بزنید (جدول ۱۸-۶).

حل

۱- میانگین دما در سالهای مشترک رابرای ایستگاه A حساب کنید.

$\bar{X}_A = 26.1$

۲- میانگین دما در سالهای مشترک را برای ایستگاه B حساب کنید.

$\bar{X}_B = 22.7$

۳- تفاضل دو مقدار $\bar{X}_A$ و $\bar{X}_B$ را به دست می‌آورید.

$\bar{X}_B - \bar{X}_A = -3.4$

جدول ۱۸-۶

سال	درجهٔ حرارت ایستگاه A	درجهٔ حرارت ایستگاه B
1352	28	-
1353	26	22
1354	30	25
1355	27	23
1356	27	24
1357	25	24
1358	24	21
1359	24	20
1360	26	-
1361	27	-
1362	28	-

۴- دما در ایستگاه B برای هریک از سالهایی که فاقد آمار است از جمع کردن رقم مشابه آن در ایستگاه A با عامل اصلاحی ۳/۴- به دست می‌آید.

$24.6 = (-3.4) + 28 =$ درجهٔ حرارت در B در سال ۱۳۵۲

$22.6 = (-3.4) + 26 =$ درجهٔ حرارت در B در سال ۱۳۶۰

$23.6 = (-3.4) + 27 =$ درجهٔ حرارت در B در سال ۱۳۶۱

$24.6 = (-3.4) + 28 =$ درجهٔ حرارت در B در سال ۱۳۶۲

۱۸-۲-۳ روش میانگین‌گیری

این روش بیشتر برای تخمین بارندگی به کار برده می‌شود. فرض می‌شود که در ایستگاه X آمار بارندگی در یک روز یا یک سال معین مفقود شده است. سه ایستگاه ۱ ، ۲ و ۳ را در مجاورت ایستگاه X در نظر می‌گیریم. اگر N_1 ، N_2 و N_3 و N_x متوسط بارندگی سالانه در ایستگاههای ۱ ، ۲ ، ۳ و X و P_1 ، P_2 ، P_3 و P_x به ترتیب آمار بارندگی (دریک‌مدت‌مشخص) درایستگاههای‌مذکور باشد، چنانچه اختلاف N_1 ، N_2 ، N_3 از N_x کمتر از ۱۰ درصد N_x باشد در این صورت برای بدست آوردن P_x از فرمول زیر استفاده می‌شود (روش میانگین‌گیری ریاضی).

$$P_x = \frac{P_1 + P_2 + P_3}{3} \qquad (۱۸-۲)$$

ولی اگر اختلاف N_1 ، N_2 و N_3 از N_x بیشتر از ۱۰ درصد N_x باشد فرمول ذیل مورد استفاده قرار می‌گیرد (روش نسبت نرمال).

$$P_x = \frac{1}{3}\left[P_1\frac{N_x}{N_1} + P_2\frac{N_x}{N_2} + P_3\frac{N_x}{N_3}\right] \qquad (۱۸-۳)$$

● مثال ۱۸-۵

ایستگاه باران‌سنجی X در طول مدت یک ماه بلااستفاده بوده است. در این مدت بارانی رخ داده است که در ایستگاه X ثبت نشده است. ولی سه ایستگاه A ، B و C که در مجاورت آن قرار دارند، این بارندگی را به ترتیب ۱۰۷، ۸۹ ، ۱۲۲ میلی‌متر نشان داده‌اند. اگر میانگین بارندگی سالانه در ایستگاههای X ، A و B و C به ترتیب ۹۷۸، ۱۱۲۰، ۹۳۵ و ۱۲۰۰ میلی‌متر باشد مقدار بارندگی در ماهی که ایستگاه X بلااستفاده بوده است چقدر تخمین زده می‌شود؟

حل

$N_x = 978$ mm

$10\%\ N_x = 97.8$ mm

$978 + 97.8 = 1075.8$ mm

$978 - 97.8 = 880.2$ mm

متوسط بارندگی سالانه ایستگاه X با ده درصد بالا و پایین ۱۰۷۵/۸ و ۸۸۰/۲ میلی‌متر است. چون متوسط بارندگی سالانه ایستگاههای A و C خارج از دو حد فوق است لذا فرمول ۱۸-۲

قابل استفاده نبوده و از فرمول ۱۸-۳ استفاده می‌کنیم.

$$P_x = \frac{1}{3}\left[P_A\left(\frac{N_x}{N_A}\right) + P_B\left(\frac{N_x}{N_B}\right) + P_C\left(\frac{N_x}{N_C}\right)\right]$$

$$P_x = \frac{1}{3}\left[107\left(\frac{978}{1120}\right) + 89\left(\frac{978}{935}\right) + 122\left(\frac{978}{1200}\right)\right]$$

$$P_x = \frac{1}{3}[93.43 + 93.09 + 99.43]$$

$$P_x = 95.3 \text{ mm}$$

اگر به جای سه ایستگاه از تعداد N ایستگاه مجاور استفاده شود فرمول تخمین باران ثبت نشده به‌صورت زیر خواهد بود.

$$\frac{P_x}{N_x} = \left(\frac{1}{N}\right)\sum_{i=1}^{N}\left(\frac{P_i}{N_i}\right) \qquad (۴-۱۸)$$

در این فرمول:

P_x = مقدار بارندگی که آمار آن موجود نبوده و باید تخمین زده شود.

N_x = متوسط سالانه بارندگی در ایستگاهی که آمار آن موجود نیست.

N = تعداد ایستگاههای اطراف که برای تخمین P_x از آنها استفاده می‌شود.

P_i = مقدار بارندگی در ایستگاه مجاور در همان زمانی که P_x ثبت نشده است.

N_i = متوسط سالانه بارندگی در ایستگاه مجاور.

روش غیر نموداری دیگر آنست که بین داده‌های مشترک بین دو ایستگاه یک رابطه همبستگی خطی یا غیر خطی برقرار کرده و معادله آن را بدست آوریم. سپس با داشتن داده‌های موجود در یک ایستگاه داده‌های غیر موجود متناظر را در ایستگاه دیگر بدست آوریم. برای این منظور می‌توان از برنامه کامپیوتری رگرسیون استفاده کرد. بطور کلی در این روش بین داده‌های موجود از یک بارش در ایستگاههای منطقه و داده‌های متوسط بارش سالانه آن یک رابطه رگرسیونی مطابق مثال زیر بدست آوده و سپس با داشتن متوسط بارش سالانه ایستگاه مورد نظر مقدار بارندگی مفقوده آن ایستگاه تخمین زده می‌شود.

● مثال ۱۸-۶

ایستگاه A در یک حوضه طی یک بارش غیرفعال بوده است حال آنکه ایستگاههای B و C و D مقدار همان بارش را به ترتیب ۱۲/۳، ۱۴/۸ و ۱۱/۹ میلی‌متر ثبت نموده‌اند. چنانچه متوسط بارش سالانه ایستگاههای A، B، C و D به ترتیب ۱۲۹ و ۱۵۱ و ۱۶۸ و ۱۳۷/۵ باشد بارش ایستگاه A در روز مورد نظر چقدر تخمین زده می‌شود.

حل

در ستون اول جدول زیر y مقدار بارش در روز مورد نظر در ایستگاههای B و C و D و در ستون دوم (X) مقادیر متوسط بارش سالانه نوشته شده است.

Y	X
12.3	151
14.8	168
11.9	137.5

با روش حداقل مربعات بین x و y رابطه ساده خطی رگرسیونی زیر را می‌توان استخراج کرد.

$$y = -1.8283 + 0.09744\,x$$

حال اگر بجای x در معادله فوق ۱۲۹ را قرار دهیم مقدار y بدست می‌آید.

$$y = -1.8283 + 0.09744\,(129) = 1.074 \text{ mm}$$

بنابرین مقدار بارش ایستگاه A در روز مورد نظر ۱۰/۷۴ میلی‌متر تخمین زده می‌شود.

۱۸-۲-۴ روش نموداری

یکی دیگر از روشهای معمول برای اصلاح داده‌های مشکوک و یا تخمین داده‌ها برای نقطه‌ای که اصولاً فاقد آمار هواشناسی یا هیدرولوژی است روش به اصطلاح گرافیکی است. مثلاً اگر بخواهیم بارندگی را برای یک نقطه تخمین بزنیم روی یک نقشه که از محل در اختیار داریم چند ایستگاه دیگر را که آمارشان تکمیل است - در همان زمانی که مورد نظر ماست - انتخاب می‌کنیم. از نقطه‌ای که بدون آمار است دو محور عمود بر هم رسم می‌کنیم. مثلاً ایستگاههای A ، B ، C ، D و E را انتخاب کرده و از نقطهٔ O که آمار آن را نداریم دو محور متعامد رسم می‌کنیم. سپس قدرمطلق مختصات هر نقطه را نسبت به این دو محور (x و y) با خط‌کش اندازه‌گیری می‌کنیم. با داشتن بارندگی در هریک از نقاط و مختصات آنها جدولی مشابه با آنچه در مثال زیر آمده است تشکیل می‌دهیم و بارندگی مورد نظر را به دست می‌آوریم.

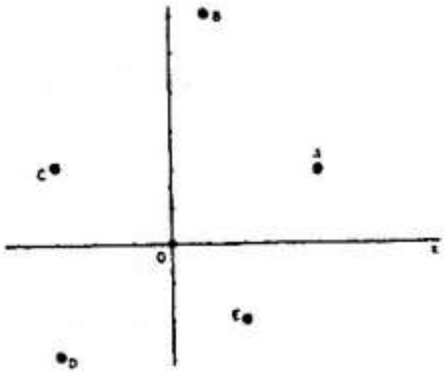
● مثال ۱۸-۷

آمار بارندگی ماه فروردین سال ۱۳۵۸ در نقطهٔ O مفقود شده است می‌خواهیم با توجه به آمار بارندگی همین ماه در ایستگاههای A ، B ، C ، D و E که روی نقشه مشخص شده است (شکل ۱۸-۴) برای این نقطه مقدار بارندگی را بدست آوریم.

حل

۱- ازنقطهٔ O دو خط عمود برهم رسم می‌کنیم تا یک دستگاه محور مختصات قائم حاصل شود.

۲- مختصات نقاط A ، B ، C ، D و E را نسبت به این دستگاه، اندازه‌گیری و مشخص کنید.



شکل ۱۸-۶ موقعیت نقطهٔ O وایستگاههای مجاور

۳- جدولی مشابه جدول ۱۸-۷ را تشکیل دهید که در آن ستون ۱ نام ایستگاهها، ستون ۲ بارندگی فروردین‌ماه در هریک از ایستگاهها، ستون ۳ مختصات طولی (x) هریک از ایستگاهها (قدرمطلق)، ستون ۴ مختصات عرضی (y) هریک از ایستگاهها (قدرمطلق)، ستون ۵ مقدار $1/(x^2 + y^2)$ برای هریک ازایستگاهها، و در ستون ۶ حاصلضرب بارندگی (ستون ۲) در $1/(x^2 + y^2)$ (ستون ۵) برای هر ایستگاه محاسبه گردیده است.

۴- میانگین بارندگی ایستگاه O در فروردین‌ماه از فرمول $P_o = \frac{\Sigma(P \times w)}{\Sigma w}$ محاسبه می‌شود. در این فرمول w همان $1/(x^2 + y^2)$ است. بطور خلاصه محاسبات جدول فوق به شرح زیر می‌باشد.

$p = mm$ بارندگی

$$w = \frac{1}{x^2 + y^2}$$

$$P.w = \frac{P}{x^2 + y^2}$$

$$P_o = \frac{\Sigma(p.w)}{\Sigma w}$$

$$P_o = \frac{5.677}{0.334}$$

$$P_o = 17 \text{ mm}$$

جدول ۱۸-۷

1	2	3	4	5	6
ایستگاه	P(mm)	x	y	w	p.w
A	16	4	2	0.05	0.80
B	18	1	6	0.027	0.486
C	15	3	2	0.077	1.154
D	20	3	3	0.055	1.112
E	17	2	2	0.125	2.125
جمع				0.334	5.677

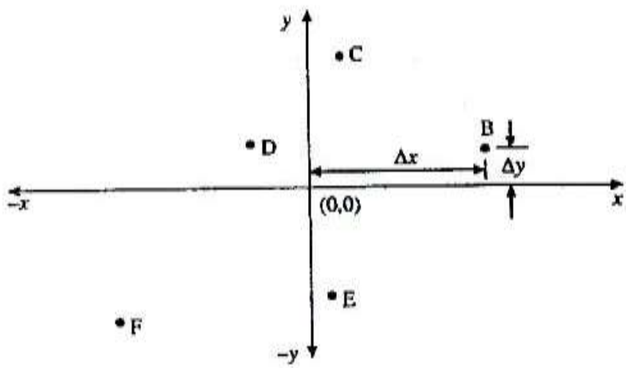
بطور کلی در این روش که بنام روش عکس فاصله (inverse distance) معروف می‌باشد برای محاسبه بارندگی مفقوده (Px) از فرمول زیر استفاده می‌شود.

$$P_x = \frac{\Sigma P_i W_i}{\Sigma W}$$

که مقدار W از معادله زیر بدست می‌آید .

$$W = \frac{1}{D^2}$$
$$D^2 = \Delta X^2 + \Delta y^2$$

مقادیر Δx و Δy در شکل زیر (۱۸-۷) نشان داده شده است.



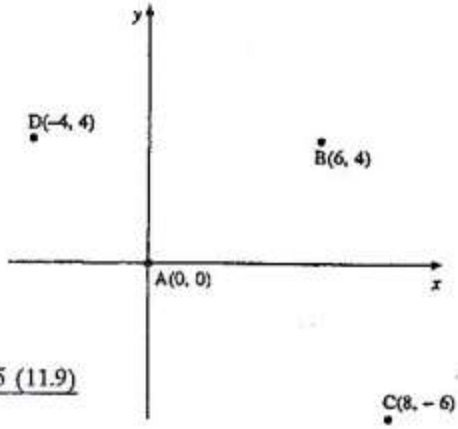
شکل ۱۸-۷

● مثال ۱۸-۸

در مثال ۱۸-۶ مرکز مختصات را ایستگاه A در نظر گرفته و ایستگاههای B و C و D به ترتیب دارای مختصات (6,4) ، (8,-6) و (4,-4) می‌باشند بارندگی مفقوده ایستگاه A چقدر است.

حل

با توجه به شکل زیر خواهیم داشت



$$D_B = 6^2 + 4^2 = 52$$

$$D_C = 8^2 + (-6)^2 = 100$$

$$D_D = (-4)^2 + (4)^2 = 32$$

$$W_B = \frac{1}{52} = 0.01923$$

$$W_C = \frac{1}{100} = 0.01$$

$$W_D = \frac{1}{32} = 0.03125$$

$$P_x = \frac{\Sigma P_i W_i}{\Sigma W_i}$$

$$P_x = \frac{0.01923\ (12.3) + 0.01\ (14.8) + 0.03125\ (11.9)}{0.01933 + 0.01 + 0.03125}$$

$$P_x = 12.5\ \text{mm}$$

۱۸-۳ آزمون کفایت داده‌ها

غالباً این سؤال مطرح می‌شود که آیا طول دوره آماری موجود برای تجزیه و تحلیل کفایت می‌کند یا خیر. مثلاً اگر از رودخانه‌ای ۱۷ سال آماری وجود داشته باشد و بخواهیم سیلهای آینده آن را پیش‌بینی نمائیم با این سؤال مواجه می‌شویم که آیا این ۱۷ سال کافی است؟ بطور کلی حداقل طول دوره آماری مورد نیاز بستگی به این دارد که بخواهیم نتایج در چه سطح آماری قابل قبول باشد. برطبق پیشنهاد ماکوس (Mackus) این حداقل از فرمول زیر محاسبه می‌شود.

$$y = [(4.30\ t)\log R]^2 + 6 \qquad (۱۸-۵)$$

در این فرمول:

y = حداقل قابل قبول تعداد داده‌ها برای تجزیه و تحلیل.

t = مقدار t استیودنت در سطح اعتماد ۹۰ درصد به ازای درجه آزادی $(y - 6)$

R = نسبت مقدار متغیر در دوره برگشت ۱۰۰ سال به مقدار آن در دوره برگشت ۲ سال بر اساس داده‌های موجود.

● **مثال ۱۸-۹**

آمار بارندگی سالانه در یک حوضه از سال ۱۳۴۴ لغایت ۱۳۶۳ به مدت ۲۰ سال به شرح زیر در اختیار می‌باشد. مقدار بارندگی را به ازای دوره‌های برگشت ۱، ۲، ۵، ۲۰ و ۱۰۰ سال

بااستفاده از قانون احتمالاتی لگاریتم نرمال محاسبه نمائید. آیا در سطح اعتماد ۹۰ درصد این‌تعداد داده‌ها برای پیش‌بینی کافی بوده است.

جدول ۸-۸

سال	بارندگی (میلی‌متر)	سال	بارندگی (میلی‌متر)
1344	371	1354	488
1345	551	1355	295
1346	307	1356	295
1347	569	1357	323
1348	594	1358	183
1349	333	1359	203
1350	488	1360	269
1351	833	1361	208
1352	284	1362	665
1353	462	1363	241

حل

با استفاده از قانون احتمال تجربی $P = \frac{m}{N+1}$ یا $T = \frac{N+1}{m}$ موقعیت مختصاتی هریک از نقاط محاسبه و روی کاغذ احتمالات لگاریتمی به شرحی که در شکل ۱۸-۸ مشاهده می‌شود رسم گردیده است (N تعداد داده‌ها می‌باشد). مطابق این شکل بارندگی به ازای دوره‌های برگشت مختلف به شرح زیر استخراج می‌شود.

جدول ۱۸-۹

٪احتمال	T(سال)	بارندگی
99	1.01	138
50	2	365
20	5	520
5	20	729
1	100	973

در فرمول ۱۸-۵ مقدار R برابر است با،

$$R = \frac{\text{باران ۱۰۰ ساله}}{\text{باران ۲ ساله}} = \frac{973}{365} = 2.67$$

لذا خواهیم داشت:

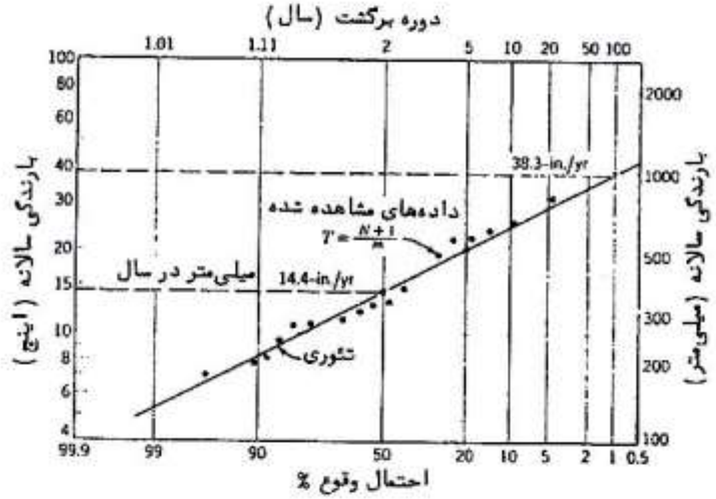
$$y = [(4.30\ t) \log R]^2 + 6$$

$$y = (4.30\ t \times 0.426)^2 + 6$$

$$y = 3.35\ t^2 + 6$$

حال برای y مقادیر مختلف را انتخاب کرده سپس از جدول t استیودنت در سطح اعتماد ۹۰

درصد و درجه آزادی y-6 مقدار t را استخراج و در معادله فوق قرار می‌دهیم. این عمل را با آزمایش و خطا آنقدر تکرار می‌کنیم تا مقادیر y و t متناسب با یکدیگر به دست آیند. در این مثال اگر y برابر ۱۷ انتخاب شود مقدار t از جدول t استیودنت برابر ۱/۷۹۶ خواهد بود که در معادله‌مذکور y برابر ۱۶/۸ به دست می‌آید و تقریباً مساوی ۱۷ می‌باشد. بنابراین برای این منظور ۱۷ سال آماری کفایت می‌کند که تعداد ۲۰ سال آمار موجود بیشتر از حداقل مورد نیاز است.



شکل ۱۸-۸

۱۸-۴ آزمون روند داده‌ها

چنانچه سری زمانی داده‌های هیدرولوژی بصورت یکنواخت سیر صعودی یا نزولی داشته باشند گوئیم که داده‌ها دارای روند (trend) می‌باشند. هر گونه تغییرات طبیعی و غیرطبیعی مانند قطع درختان جنگل، احداث شهرک‌ها و یا زمین‌لغزه‌های بزرگ باعث تغییر در روند داده‌ها می‌شوند. برای آنکه وجود یا عدم وجود روند در داده‌ها را آزمایش کنیم سه روش عمده بکار می‌رود که عبارتند از:

- آزمایش نقاط عطف یا چرخش (turning point test)
- آزمون کندال (Kendal)
- آزمون رگرسیون خطی

۱۸-۴-۱ آزمون نقاط چرخش

الف- کلیه داده‌ها را به ترتیب زمان وقوع ردیف کنید (فرضاً N عدد)

ب- تعداد نقاط چرخش را درسری داده‌ها بدست آورید. نقطه چرخش به حالتی گفته

می‌شودکه هر عدد هم از عدد ماقبل و هم از مابعد خود بزرگتر باشد $[x_{i-1}<x_i>x_{i+1}]$ و یا آنکه هر عدد هم از عدد ما قبل خود و هم از عدد ما بعد خود کوچکتر باشد $[x_{i-1}>x_i<x_{i+1}]$.

ج - فرض کنید تعداد نقاط چرخش P باشد

د - تعداد نقاط چرخش مورد انتظار(E(P را از فرمول زیر بدست آورید.

$$E(P) = \frac{2\,(N - 2)}{3} \qquad (۶-۱۸)$$

هـ - واریانس P را [(Var (P] از فرمول زیر بدست می‌آورید

$$Var(P)=\frac{16N - 29}{90} \qquad (۷-۱۸)$$

و- مقدار z از فرمول زیر محاسبه کنید. z توصیف کننده p بر حسب استاندارد نرمال است.

$$z = \frac{[P - E(P)]}{[Var\ (P)]^{0.5}} \qquad (۸-۱۸)$$

ز - مقدار z را در سطح معنی‌دار بودن ۵ درصد آزمایش کنید. چنانچه z کوچکتر از ۱/۹۶+ و بزرگتر از ۱/۹۶- باشد خواهیم گفت که داده‌ها تصادفی بوده و فاقد روند می‌باشند.

● مثال ۱۸-۱۰

طی ۱۰ سال آمارگیری از مقدار بارندگی سالانه در یک ایستگاه اعداد زیر بدست آمده است. آیا داده‌ها دارای روند می‌باشند یا خیر.

جدول ۱۸-۱۰

سال	۱۹۹۱	۱۹۹۲	۱۹۹۳	۱۹۹۴	۱۹۹۵	۱۹۹۶	۱۹۹۷	۱۹۹۸	۱۹۹۹	۲۰۰۰
بارندگی (mm)	۶۵۲	۱۵۳	۷۳۲	۵۰۰	۳۵۷	۴۹۹	۶۷۶	۵۲۶	۴۸۲	۵۱۴

حل

با توجه به داده‌های بارندگی جدول ۱۸-۱۱ را تشکیل می‌دهیم.

جدول ۱۸-۱۱

سال	داده‌ها	آزمون برای نقاط عطف			وجودنقطه عطف	تعداد تجمعی نقاط عطف
(1)	(2)	(3)			(4)	(5)
1991	652	-	-	-		
1992	153	652	153	732	1	1
1993	732	153	732	500	1	2
1994	500	732	500	357	-	-
1995	357	500	357	499	1	3
1996	499	357	499	676	-	-
1997	676	499	676	526	1	4
1998	526	676	526	482	-	-
1999	482	526	482	514	1	5
2000	514	-	-	-	-	

در جدول مشاهده می‌شود که مثلاً ۱۵۳ (ستون ۳ ردیف ۲) از هر دو عدد مجاور کوچکتر است، پس این یک نقطه چرخش است. در ردیف دوم ۷۳۲ نیز از هر دو عدد مجاور بزرگتر است لذا این نقطه نیز یک نقطه چرخش است. به همین ترتیب مشاهده می‌شود که در بین داده‌ها ۵ نقطه چرخش وجود دارد.

تعداد داده‌ها $N = 10$

تعداد نقاط چرخش $P = 5$ (با توجه به اعداد جدول)

$$E(P) = \frac{2(N - 2)}{3} = \frac{2(10 - 2)}{3} = 5.33$$

$$Var(P) = \frac{(16N - 29)}{90} = \frac{16 \times 10 - 29}{90} = 1.455$$

$$Z = \frac{P - E(P)}{[Var(P)]^{0.5}} = \frac{5 - 5.33}{(1.455)^{0.5}} = - 0.276$$

چون ۰/۲۷۶- بین ۱/۹۶+ و ۱/۹۶- می‌باشد لذا در سطح اعتماد ۵ درصد می‌توان گفت که در داده‌ها روند وجود ندارد.

۱۸-۴-۲ آزمون کندال

الف- داده‌ها را به ترتیب وقوع ردیف کنید. فرض کنید تعداد آنها n باشد.

ب - اولین داده (x_1) را گرفته و به ترتیب آن را با سایر داده‌ها مقایسه کنید و تعیین کنید در چند مورد از دیگر داده‌ها بزرگتر است و تعداد آن را P_{1ex} بنامید.

ج - دومین داده (x_2) را گرفته و آن را با داده‌های بعد از خود مقایسه کنید و مشخص کنید در چند مورد از دیگر داده‌ها بزرگتر است و تعداد آن را P_{2ex} بنامید.

د- عمل فوق را برای داده‌های دیگر ($x_3, x_4, ... x_n$) آزمایش کرده و مقادیر P_{3ex}، P_{4ex}،... و P_{nex} را بدست آورید.

ه - جمع مقادیر P_{1ex}، P_{2ex}،... و P_{nex} را بدست آورید و آن را P بنامید.

$$P = P_{1ex} + P_{2ex} + P_{3ex} + ... + P_{nex} \qquad (۱۸-۹)$$

و - مقدار E(P) را از فرمول زیر بدست آورید.

$$E(P) = \frac{n(n - 1)}{4} \qquad (۱۸-۱۰)$$

ز- ضریب کندال (τ) را از فرمول زیر محاسبه کنید:

$$\tau = \left[\left\{\frac{4P}{n(n - 1)}\right\}-1\right] \qquad (۱۸-۱۱)$$

ح - واریانس τ را از فرمول زیر محاسبه کنید.

$$Var(\tau) = \left[\frac{\{2(2n + 5)\}}{9n(n + 1)}\right] \qquad (۱۸-۱۲)$$

ط – مقدار z را از فرمول زیر بدست آورید

$$z = \frac{\tau}{[\mathrm{Var}(\tau)]^{0.5}} \qquad (۱۳-۱۸)$$

ی – چنانچه z بزرگتر از ۱/۹۶+ و یا کوچکتر از ۱/۹۶- باشد خواهیم گفت داده‌ها دارای روند می‌باشند. در غیر این صورت داده‌ها تصادفی و بدون روند است.

● مثال ۱۸-۱۱

مقدار بارندگی سالانه در یک حوضه طی سال‌های ۱۹۹۱ تا ۲۰۰۰ به ترتیب به شرح زیر مثال ۱۸-۱۰ می‌باشد. با آزمون کندال وجود روند در داده‌ها را آزمایش کنید.

حل

با مقایسه اولین داده (۶۲۵) با سایر داده‌ها مشاهده می‌شود که در ۷ مورد بزرگتر از دیگر داده‌هاست ($P_{1ex} = 7$). با مقایسه داده دوم، سوم و .... مشاهده خواهد شد که مقادیر P_{2ex} و P_{3ex} و ... چقدر بوده که به ترتیب جدول زیر آورده شده‌اند.

ردیف	داده	P_{ex}
1	652	$P_{1ex}=7$
2	153	$P_{2ex}=0$
3	732	$P_{3ex}=7$
4	500	$P_{4ex}=4$
5	357	$P_{5ex}=0$
6	499	$P_{6ex}=1$
7	676	$P_{7ex}=3$
8	526	$P_{8ex}=2$
9	482	$P_{9ex}=0$
10	514	-
جمع		23

$n = 10$

$p = 23$

$$E(P) = \frac{n(n-1)}{4} = \frac{10(10-1)}{4} = 22.5$$

$$\tau = \left[\left\{\frac{4P}{n(n-1)}\right\} - 1\right] = \left\{\frac{4 \times 23}{(10\times 9)} - 1\right\} = 0.022$$

$$\mathrm{Var}(\tau) = \left[\frac{\{2(2n+5)\}}{9n(n-1)}\right] = \frac{2(2 \times 10 + 5)}{[9 \times 10(10-1)]} = 0.0617$$

$$z = \frac{\tau}{[Var(\tau)]^{0.5}} = \frac{0.022}{(0.0617)^{0.5}} = 0.0894$$

چون z از ۱/۹۶- بزرگتر و از ۱/۹۶+ کوچکتر است لذا در سطح ۵ درصد داده‌ها تصادفی بوده و فاقد روند می‌باشند.

۱۸-۴-۳ رگرسیون خطی

در این روش در یک دستگاه مختصات در محور افقی سال و در محور عمودی مقادیر داده‌های مربوط به آن سال آورده می‌شود و یک خط مستقیم با معادله $y = a + bx$ از بین نقاط گذرانده می‌شود و سپس درجه همبستگی آنها به روش‌های آماری مشخص می‌گردد. توجه شود که این روش فقط برای سری سالانه بکار می‌رود.

۱۸-۴-۴ تخمین و حذف روند

برای کمی کردن و حذف هر گونه روند یا نوسانات در سری داده‌های هیدرولوژیکی دو روش معمول است که عبارتند از: روش میانگین متحرک و روش حداقل مربعات.

الف - میانگین متحرک

در روش موسوم به میانگین متحرک نوسانات کوچک داده‌های هیدرولوژیکی صاف می‌شوند. فرض کنید داده‌های بارندگی (x_i) به مدت ۱۲ سال به شرح جدول ۱۸-۱۲ در اختیار می‌باشد. اگر این داده‌ها را نسبت به زمان در یک دستگاه محور مختصات رسم کنیم (شکل ۱۸-۹ الف) مشاهده خواهد شد که این نمودار دارای نوسانات زیادی است. حال اگر سه رقم اول یعنی x_1 و x_2 و x_3 را که مربوط به سال‌های ۱۹۵۰، ۱۹۵۱ و ۱۹۵۳ می‌باشد میانگین گیری کرده و عدد بدست آمده را به جای x_2 قرار دهیم عددی که بدست می‌آید (x_2) برابر ۲۷۱/۶ خواهد شد.

$x_2 = 273$

$$x_2' = \frac{x_1 + x_2 + x_3}{3} = \frac{251 + 2731 + 291}{3} = 271.6$$

لذا عدد اول (سال ۱۹۵۰ با بارندگی ۲۵۱ میلی‌متر) خود بخود حذف شده و به جای عدد دوم در سال ۱۹۵۱ که ۲۷۳ بود رقم ۲۷۱/۶ قرار می‌دهیم به همین طریق میانگین x_2 و x_3 و x_4 را گرفته و به جای ۲۹۱ که مربوط به سال ۱۹۵۲ یا x_3 بود قرار می‌دهیم.

$x_3 = 291$

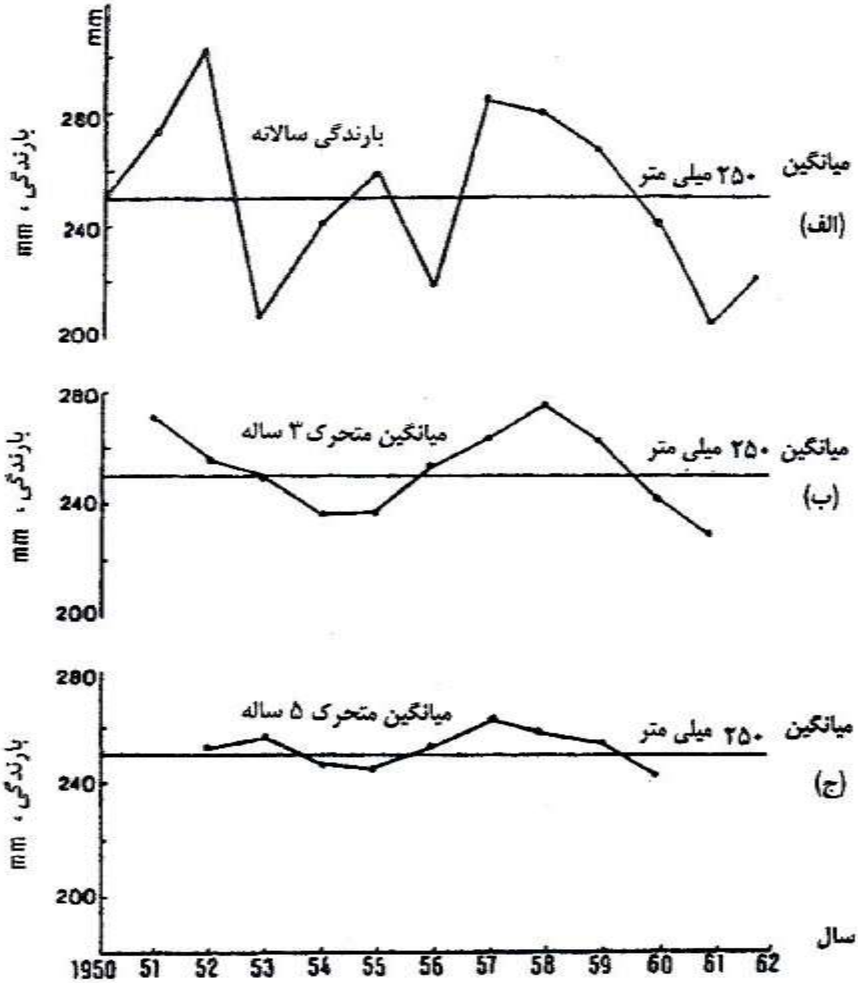
$$x_3 = \frac{x_1 + x_3 + x_4}{3} = \frac{273 + 291 + 214}{3} = 259.3$$

این عمل را به همین طریق تکرار کرده و میانگین سال‌های ۱۹۵۲، ۱۹۵۳ و ۱۹۵۴ را محاسبه کرده و عدد حاصله را به جای بارندگی در سال ۱۹۵۳ قرار می‌دهیم و الی آخر. در جدول ۱۸-۱۲ ستون سوم نتیجه این محاسبات را نشان میدهد. ملاحظه می‌شود چون ما برای صاف کردن

داده‌ها از سه عدد استفاده کرده تعداد داده‌های ما در سری آمار به اندازه ۲ عدد (اول و آخر) حذف می‌شود. چنانچه به جای ۳ سال ۵ سال را ملاک قرار می‌دادیم و میانگین سال‌های ۱۹۵۰ تا ۱۹۵۴ را محاسبه می‌کردیم و برای سال ۱۹۵۲ منظور می‌کردیم ۲ رقم از اول سری و ۲ رقم از آخر سری داده‌ها از دست می‌دادیم (ستون چهارم جدول ۱۸-۱۲). رسم داده‌های مربوط به میانگین متحرم ۳ و ۵ ساله در شکل ۱۸-۹ ب و ۱۸-۹ ج نشان می‌دهد که نوسانات موجود در سری داده‌ها تا حد زیادی از بین رفته است. بدین ترتیب که در سری داده‌های اولیه دامنه تغییرات ۸۱ میلی‌متر (از حداکثر ۲۹۱ تا حداقل ۲۱۰ میلی‌متر) در داده‌های صاف شده با میانگین متحرک ۳ ساله این دامنه ۴۷/۶ میلی‌متر (از ۲۷۵/۶ تا ۲۸۸) و در میانگین متحرک ۵ ساله این دامنه به ۱۸ میلی‌متر (حداکثر ۲۶۱ و حداقل ۲۴۳) تقلیل یافته است. باید توجه داشت که اگر طول مدت آماری بسیار زیاد و بیشتر از ۱۰۰ سال باشد می‌توان میانگین متحرک ۵ و ۷ و ۹ ساله را نیز اعمال کرد ولی برای سری‌های آماری کوتاه صاف کردن داده‌ها با میانگین ۳ ساله کفایت می‌کند.

جدول ۱۸-۱۲

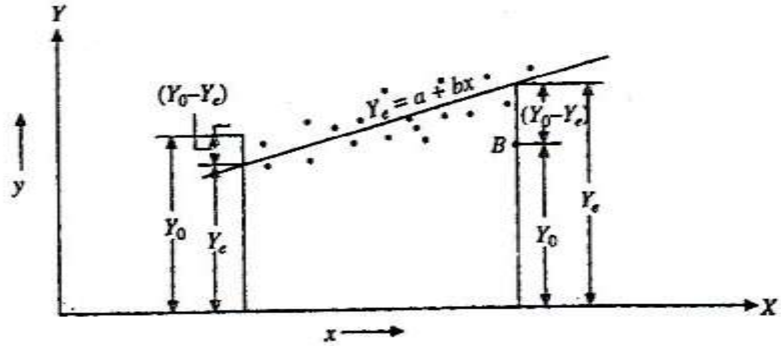
سال	بارندگی (mm)	میانگین متحرک ۳ ساله	میانگین متحرک ۵ ساله
1950	251	-	-
1951	273	271.6	-
1952	291	259.3	254
1953	214	249.0	255
1954	242	237.0	245
1955	257	240.0	243
1956	221	253.3	256
1957	282	261.6	261
1958	280	275.6	258
1959	265	262.3	255
1960	242	239.0	245
1961	210	228.0	-
1962	232	-	-



شکل ۱۸-۹ میانگین متحرک ۳ ساله

ب - حداقل مربعات

چنانچه یک سری داده‌های هیدرولوژیکی (y) را به صورت سری در سال‌های مختلف داشته باشیم و آن‌ها را در یک دستگاه محور مختصات نسبت به زمان (x) رسم کنیم می‌توانیم از بین آن‌ها یک خط مستقیم عبور داده و به روش حداقل مربعات مطابق شکل ۱۸-۱۰ معادله خط بهترین برازش را بدست آوریم. $(y_c = a + bx)$. چنانچه مقدار b زیاد باشد نشان دهنده آن است که داده‌های دارای نوسانات است. حال می‌توانیم برای هر سال مقدار مشاهده شده را که در سری داده‌ها موجود است با مقدار متناظری که ازمعادله خط بدست می‌آید مقایسه کرده وتفاوت آن‌ها $(y_o - y_c)$ را به داده اولیه افزوده و یا از آن کسر کرد تا نوسانات داده‌ها از بین برود.



شکل ۱۸-۱۰

مسائل

۱۸-۱ داده‌های مربوط به حداکثر بارندگی ۲۴ ساعته در سال در ایستگاهی واقع در یک حوضهٔ آبریزبه شرح زیر در دست است. با روش «ران تست» تصادفی بودن داده‌ها را بررسی نمایید، ۵۰،۵۵، ۵۷، ۳۷، ۵۴، ۳۴، ۳۱، ۴۵، ۴۵، ۵۱، ۶۷، ۲۹، ۷۶، ۶۲، ۴۴، ۵۴، ۵۶، ۳۵، ۵۰، ۲۷.

۱۸-۲ در یک حوضه آبریز ایستگاه A در زمان وقوع بارندگی بلا استفاده بوده است حال آنکه در همین زمان ایستگاههای B و C و D که در مجاورت ایستگاه A بوده‌اند بارندگی را ثبت کرده‌اند. که مقادیر بارش در آنها ۱۲۳، ۱۴۸ و ۱۱۹ میلی‌متر می‌باشد میانگین بارندگی سالانه در ایستگاههای A و B و C و D به ترتیب ۱۲۹۰، ۱۵۱۰، ۱۶۸۰، و ۱۳۷۹ میلی‌متر است حساب کنید بارندگی ثبت نشده ایستگاه را با روش هایی که می‌دانید. محاسبه کنید.

جواب: روش میانگین‌گیری ریاضی صادق نیست، با روش نسبت‌های نرمال (۸۲/۶ میلی‌متر)، با روش رگرسیون خطی (۱۰۷/۴).

۱۸-۳ در یک حوضهٔ آبریز بارندگی سالانهٔ ایستگاه X از سال ۱۹۳۷ لغایت ۱۹۷۲ به شرح جدول ۱۸-۱۳ در اختیار است. میانگین بارندگی ایستگاههای دیگر حوضه نیز در سالهای مشابه محاسبه شده و در ستون سوم جدول مذکور نوشته شده است به روش جرم مضاعف همگنی داده‌های ایستگاه X را بررسی کنید.

جدول ۱۸-۱۳

سال	بارندگی سالانه (mm)		سال	بارندگی سالانه (mm)	
	ایستگاه X	متوسط منطقه		ایستگاه X	متوسط منطقه
1972	188	264	1954	223	360
1971	185	228	1953	173	234
1970	310	386	1952	282	333
1969	295	297	1951	218	236
1968	208	284	1950	246	251
1967	287	350	1949	284	284
1966	183	236	1948	493	361
1965	304	371	1947	320	282
1964	228	234	1946	274	252
1963	216	290	1945	322	274
1962	224	282	1944	437	302
1961	203	246	1943	389	350
1960	284	264	1942	305	228
1959	295	332	1941	320	312
1958	206	231	1940	328	284
1957	269	234	1939	308	315
1956	241	231	1938	302	280
1955	284	312	1937	414	343

۱۸-۴ به روش نسبت‌ها و تفاضل‌ها داده‌های مفقود شده در ایستگاه x در جدول ۱۸-۱۴ را تکمیل کنید.

جدول ۱۸-۹

سال	ایستگاه مبنا		ایستگاه X	
	بارندگی سالانه	درجه حرارت سالانه	بارندگی سالانه	درجه حرارت سالانه
1350	286	12	380	14
1351	290	11	350	17
1352	350	10	315	-
1353	180	14	271	-
1354	210	13	-	12
1355	220	11	-	17
1356	310	12	320	15
1357	300	15	270	14
1358	270	10	243	11
1359	242	18	262	11

۱۸-۵ در یک دستگاه محور مختصات میانگین متحرک ۳ ساله و ۵ ساله متوسط بارندگی سالانه ایستگاههای حوضه را (ستون ۳ جدول ۱۸-۸) رسم کنید.

منابع برای مطالعه بیشتر

2- Chow, V.T., *Hand book of applied hydrology,* McGraw Hill, New York, 1964.

3- Linsley, R. et al, *Hydrology for engineers,* McGraw Hill, New York, 1958.

4- Raudkivi, A, *Hydrology, pergamon press,* Oxford, 1979.

5- Schwab, G. et al, *Soil and water conservation engineering,* John Wiley and Sons, New York, 1981.

6- Shaw, E., *Hydrology in practice,* Van Nostrand Reinhold, London, 1988.

7- Viessman, W. et al, *Introduction to hydrology,* IEP, New York, 1972.

8- Wilson, E., *Engineering hydrology,* Mac Millan, London, 1984.