

# آمار کاربردی برای مهندسان عمران و محیط زیست (جلد اول)

مترجمان:

محمد مهدی باطنی، جواد بهمنش

# نمایش کامل رزومه



کوکوژا، ارنزو Kocoujo, A. T	سرسامه
آمار کاربردی برای مهندسان عمران و محیط زیست، انتشارات دانشگاه ب. کوکوژا و ریزو روس - شرکتان متحد آموی ناقص، خود نهایی.	عنوان و نامردی ناظر
رومیه: دانشگاه ارومیه، ۱۳۹۴	مشخصات نشر
ع.ت. حضور، جدول، محور: ۲۲ x ۲۹ سیم	مشخصات ناشر
انتشارات دانشگاه ارومیه، ۱۳۹۴	فرونت
978-600-6648-36-3	پشتاک
کتاب	وضعیت فهرست نویسی
Applied statistics for civil and environmental engineers... عنوان اصلی	زادگاه
آمار ریاضی - ریاضیات آماری (عالی)	موضوع
آمار ریاضی - مسائل آماری ها و غیره (۱۳۴)	نوع و
مهندسی عمران - الگوهای ریاضی	موضوع
مهندسی محیط زیست	موضوع
رومیه، روسو	سیاسته افروده
Russo, Renzo	سیاسته افروده
ناقص، خود نهایی، ۱۳۹۴ - ترجم	سیاسته افروده
نویسن، مواد، ۱۳۵۲ - ترجم	سیاسته افروده
دانشگاه ارومیه	رده بک کنگره
QATV3/107,51/0, 1197	رده بک دولی
01900-97	شماره کتابشناسی ملی
91124227	

تاریخ درخواست : 1397/12/19  
تاریخ پاسخگویی :  
کد بکری : 4148439

ارسال

ارسال فرمهای آماده به نشانی الکترونیکی: [lib@guilan.ac.ir](mailto:lib@guilan.ac.ir) یا [lib@guilan.ac.ir](mailto:lib@guilan.ac.ir)

صفحه	عنوان
2	فهرست مطالب
7	فصل یکم: تجزیه و تحلیل اولیه داده ها
7	1.1 نمایش نموداری
8	1.1.1 نمودار خطی یا نمودار میله ای
8	1.1.2 نمودار نقطه ای
9	1.1.3 هیستوگرام
13	1.1.4 چند ضلعی های فراوانی
16	1.1.5 نمودار فراوانی نسبی تجمعی
	1.1.6 منحنی مدت
	1.1.7 خلاصه ای از بخش 1.1
	1.2 خلاصه عددی داده ها
	1.2.1 شاخص های تمایل مرکزی
22	1.2.2 شاخص های پراکندگی
	1.2.3 شاخص های عدم تقارن
	1.2.4 شاخص های قلگی
	1.2.5 خلاصه ای از بخش 1.2
	1.3 روش های کاوشگرانه
	1.3.1 نمودار ساقه و برگ
	1.3.2 نمودار جعبه ای
	1.3.3 خلاصه ای از بخش 1.3
	1.4 داده های مشاهده شده به صورت جفت
	1.4.1 همبستگی و نمودارهای گرافیکی
	1.4.2 کوواریانس و ضریب همبستگی
	1.4.3 نمودارهای Q-Q
	1.4.4 خلاصه قسمت 1.4
	1.5 خلاصه فصل اول
	<b>Error! Bookmark not defined.</b>
	<b>Error! Bookmark not defined.</b> منابع
	<b>Error! Bookmark not defined.</b> مسائل
	<b>Error! Bookmark not defined.</b> فصل دوم: مفاهیم پایه احتمال
	2.1 وقایع تصادفی
	2.1.1 فضای نمونه و پیشامد
	2.1.2 پیشامد تهی، اشتراک و اجتماع
	2.1.3 نمودار ون و فضای پیشامد
	2.1.4 خلاصه قسمت 2.1
	2.2 معیارهای احتمال
	<b>Error! Bookmark not defined.</b>
	2.2.1 تفسیر احتمال
	2.2.2 اصول احتمال
	2.2.3 قاعده جمع
	2.2.4 دیگر خواص توابع احتمال
	2.2.5 احتمال شرطی و قاعده ضرب
	2.2.6 استقلال تصادفی
	2.2.7 استقلال احتمال کل و تئوری بیز
	2.2.8 خلاصه قسمت 2.2
	2.3 خلاصه فصل 2
	<b>Error! Bookmark not defined.</b> منابع
	<b>Error! Bookmark not defined.</b> مسائل
	<b>Error! Bookmark not defined.</b> فصل سوم: متغیرهای تصادفی و خواص آنها
	3.1 متغیرهای تصادفی و توزیعهای احتمال
	3.1.1 متغیرهای تصادفی
	3.1.2 تابع جرم احتمال (pmf)

3.1.3	تابع توزیع تجمعی یک متغیر تصادفی گسسته
3.1.4	تابع چگالی احتمال
3.1.5	تابع توزیع تجمعی یک متغیر تصادفی پیوسته
3.1.6	خلاصه قسمت 3.1
3.2	توصیف متغیرهای تصادفی
3.2.1	امید ریاضی و دیگر شاخص های جامعه
3.2.2	توابع مولد
3.2.3	برآورد پارامترها
3.2.4	خلاصه ای از بخش 3.2
3.3	متغیرهای تصادفی چندگانه
3.3.1	توزیع های احتمال توام برای متغیرهای گسسته
3.3.2	توزیع احتمال توام متغیرهای پیوسته
3.3.3	خواص متغیرهای چندگانه
3.3.4	خلاصه ای از بخش 3.3
3.4	متغیرهای تصادفی مرتبط و احتمالات آن
3.4.1	توابعی از یک متغیر تصادفی
3.4.2	توابعی از دو یا چند متغیر تصادفی
3.4.3	خواص متغیرهای مشتق شده
3.4.4	متغیرهای مرکب
3.4.5	خلاصه ای از بخش 3.4
3.5	کوپلا (مفصل)
3.6	خلاصه فصل 3
	منابع
	مسائل
	فصل چهارم: توزیع های احتمال
4.1	توزیع های گسسته
4.1.1	توزیع برنولی
4.1.2	توزیع دو جمله ای
4.1.3	توزیع پواسون
4.1.4	توزیع های دو جمله ای منفی و هندسی
4.1.5	توزیع سری لگاریتمی
4.1.6	توزیع چند جمله ای
4.1.7	توزیع فوق هندسی
4.1.7	خلاصه ای از بخش 4.1
4.2	توزیع های پیوسته
<b>Error! Bookmark not defined.</b>	
4.2.1	توزیع یکنواخت
4.2.2	توزیع نمایی
4.2.3	توزیع ارلانگ و گاما
4.2.4	توزیع بتا
4.2.5	توزیع وایبل
4.2.6	توزیع نرمال
4.2.7	توزیع لوگ نرمال
4.2.8	خلاصه ای از بخش 4.2
4.3	توزیع های چند متغیره
4.3.1	توزیع نرمال دو متغیره
4.3.2	سایر توزیع های دو متغیره
4.4	خلاصه فصل 4
	منابع
	مسائل
	فصل پنجم: برآورد مدل و آزمون سازی
5.1	مروری بر اصطلاحات مربوط به نمونه گیری تصادفی
5.2	خواص برآوردگرها
5.2.1	نارایی
5.2.2	سازگاری
5.2.3	کمینه واریانس

Error! Bookmark not defined.....	5.2.4 کارایی
Error! Bookmark not defined.....	5.2.5 بسندگی
Error! Bookmark not defined.....	5.2.6 خلاصه بخش 5.2
<b>Error! Bookmark not defined.....</b>	<b>5.3 برآورد بازه اطمینان</b>
Error! Bookmark not defined.....	5.3.1 برآورد بازه اطمینان میانگین هنگامی که انحراف معیار معلوم است
Error! Bookmark not defined.....	5.3.2 برآورد بازه اطمینان برای میانگین هنگامی که انحراف معیار مجهول است
Error! Bookmark not defined.....	5.3.3 برآورد بازه اطمینان برای یک نسبت
Error! Bookmark not defined.....	5.3.4 توزیع نمونه گیری تفاضل ها و مجموع های آماره ها
Error! Bookmark not defined.....	5.3.5 برآورد بازه برای واریانس: توزیع مربع کای
Error! Bookmark not defined.....	5.3.6 خلاصه بخش 5.3
<b>Error! Bookmark not defined.....</b>	<b>5.4 آزمون فرض</b>
Error! Bookmark not defined.....	5.4.1 رویه آزمون
Error! Bookmark not defined.....	5.4.2 احتمالات خطاهای نوع I و نوع II و تابع توان
Error! Bookmark not defined.....	5.4.3 لم نیمن- پیرسون
Error! Bookmark not defined.....	5.4.4 آزمون های فرض مربوط به واریانس
Error! Bookmark not defined.....	5.4.5 توزیع $F$ و کاربرد آن
Error! Bookmark not defined.....	5.4.6 خلاصه قسمت 5.4
<b>Error! Bookmark not defined.....</b>	<b>5.5 روش های ناپارامتری</b>
Error! Bookmark not defined.....	5.5.1 آزمون علامت برای میانه
Error! Bookmark not defined.....	5.5.2 آزمون رتبه ای علامتدار ویلکاکسون برای پیوند مشاهدات جفت شده
Error! Bookmark not defined.....	5.5.3 آزمون کروسکال-والیس برای مشاهدات جفت شده در $k$ نمونه
Error! Bookmark not defined.....	5.5.4 آزمون های تصادفی بودن: آزمون گردش
Error! Bookmark not defined.....	5.5.5 ضریب همبستگی رتبه ای اسپیرمن
Error! Bookmark not defined.....	5.5.6 خلاصه بخش 5.5
<b>Error! Bookmark not defined.....</b>	<b>5.6 آزمون های نیکویی برآزش</b>
Error! Bookmark not defined.....	5.6.1 آزمون نیکویی برآزش مربع کای
Error! Bookmark not defined.....	5.6.2 آزمون نیکویی برآزش کولموگروف اسمیرنوف
Error! Bookmark not defined.....	5.6.3 آزمون کولموگروف اسمیرنوف دو نمونه ای
Error! Bookmark not defined.....	5.6.4 آزمون نیکویی برآزش اندرسون-دارلینگ
Error! Bookmark not defined.....	5.6.5 روش های دیگر آزمون نیکویی برآزش برای توزیع نرمال
Error! Bookmark not defined.....	5.6.6 خلاصه بخش 5.6
<b>Error! Bookmark not defined.....</b>	<b>5.7 تحلیل واریانس</b>
Error! Bookmark not defined.....	5.7.1 تحلیل واریانس یک طرفه
Error! Bookmark not defined.....	5.7.2 تحلیل واریانس دو طرفه
Error! Bookmark not defined.....	5.7.3 خلاصه بخش 5.7
<b>Error! Bookmark not defined.....</b>	<b>5.8 روش های رسم نمودار احتمال و ابزار کمی بصری</b>
Error! Bookmark not defined.....	5.8.1 رسم نمودار احتمال برای توزیع یکنواخت
Error! Bookmark not defined.....	5.8.2 رسم نمودار احتمال برای توزیع نرمال
Error! Bookmark not defined.....	5.8.3 رسم نمودار احتمال برای توزیع گامبل یا EV1
Error! Bookmark not defined.....	5.8.4 رسم نمودار احتمال برای توزیع های دیگر
Error! Bookmark not defined.....	5.8.5 روش های بصری برآزش بر اساس هیستوگرام
Error! Bookmark not defined.....	5.8.6 خلاصه بخش 5.8
<b>Error! Bookmark not defined.....</b>	<b>5.9 شناسایی و تطابق دادن داده های پرت</b>
Error! Bookmark not defined.....	5.9.1 آزمونهای فرض
Error! Bookmark not defined.....	5.9.2 آماره های آزمون برای تشخیص داده های پرت
Error! Bookmark not defined.....	5.9.3 کار با داده های غیرنرمال
Error! Bookmark not defined.....	5.9.4 برآورد احتمال پیشامدهای حدی زمانی که داده های پرت وجود دارند
Error! Bookmark not defined.....	5.9.5 خلاصه بخش 5.9
.....	5.10 خلاصه فصل 5
<b>Error! Bookmark not defined.</b>	
Error! Bookmark not defined.....	منابع
Error! Bookmark not defined.....	مسائل
Error! Bookmark not defined.....	پیوست ها
<b>Error! Bookmark not defined.....</b>	<b>1. پیوست الف</b>
<b>Error! Bookmark not defined.....</b>	<b>2. پیوست ب</b>
<b>Error! Bookmark not defined.....</b>	<b>3. پیوست پ</b>
<b>Error! Bookmark not defined.....</b>	<b>4. پیوست ت</b>



## پیش گفتار مترجمان:

مهندسان عمران و محیط زیست می‌بایست درک درستی از آمار ریاضی و نظریه احتمال برای مواجهه با تغییرپذیری و تنوعی که در مسائل مختلف این حوزه وجود دارد، داشته باشند. از مهمترین این مسائل تغییرات اقلیمی هستند که طیف وسیعی از رخدادهای احتمالی را با احتمال وقوع متفاوت شامل می‌شوند.

علم آمار و احتمال یک ابزار مهم در طبقه بندی اطلاعات مربوط به فعالیت‌های انسانی موثر بر سامانه آب و هوایی و همچنین اطلاعات مربوط به چگونگی تاثیر تغییرات اقلیمی بر سامانه‌های طبیعی و انسانی ارائه می‌کند. از سوی دیگر، مطابق با گزارش کمیته‌های تخصصی بین‌المللی تغییرات اقلیمی اثرات شگرفی بر آب و هوای منطقه خاورمیانه و به خصوص ایران داشته و خواهد داشت. لذا لازم است که فعالان این حوزه در کشور آشنایی کافی با مفاهیم آمار کاربردی داشته باشند.

کتاب حاضر قابل استفاده برای همه دانشجویان و مهندسان عمران و محیط زیست بخصوص پژوهشگران حوزه منابع آب و تغییرات اقلیمی است. دانشجویان در مقاطع تحصیلات تکمیلی و پایین‌تر به فراخور خود می‌توانند از این کتاب بهره ببرند. موضوعات متداول در این حوزه به یک شیوه روشن و به روز و با استفاده از رویکردی فرآیند محور مطرح شده است. ۳۲۵ مثال و مطالعه موردی از اروپا و آمریکا که همگی در زمینه‌های مختلف کاری مهندسان عمران و محیط زیست است در کتاب آورده شده و هر فصل نیز در انتهای آن دارای مسائل واقعی برای حل کردن است.

جلد نخست این کتاب شامل مباحث مقدماتی نظیر تحلیل اولیه داده‌ها، مفاهیم پایه‌ای احتمال، متغیرهای تصادفی و خصوصیات آن‌ها، توزیع‌های متداول احتمالی و برآورد مدل و آزمون‌های آن است. در جلد دوم به عناوین پیشرفته‌تری چون رگرسیون و تحلیل چند متغیره، تحلیل فراوانی مقادیر حدی، روش‌های شبیه سازی به منظور طراحی، تحلیل ریسک و قابلیت اطمینان و روشهای بیزی تصمیم سازی و عدم قطعیت پارامترها پرداخته می‌شود.

بدون شک این برگردان دارای کمبودها، اشتباهات و کاستی‌هایی است که با تذکر همکاران و دانشجویان عزیز سعی خواهد شد در چاپ‌های بعدی اصلاح گردد. لازم می‌دانیم از مسئولان محترم دانشگاه ارومیه جهت فراهم آوردن امکانات لازم برای چاپ کتاب سپاسگزار می‌نماییم. در پایان از ویراستاران علمی این اثر آقایان دکتر یعقوب دین پژوه، دکتر حسین رضایی و دکتر مجید منتصری به خاطر راهنمایی‌های ارزندشان و کمک همه عزیزانی که ما را در ارائه این اثر یاری داده‌اند قدر دانی می‌شود.

جواد بهمنش

j.behmanesh@urmia.ac.ir

محمد مهدی باطنی

mm.bateni@urmia.ac.ir

۱۳۹۴ شمسی

## فصل یکم: تجزیه و تحلیل اولیه داده‌ها

می‌توان مشاهده کرد که مرکز داده‌ها نزدیک به 40 نیوتن بر میلی‌متر مربع است، اما گسترده‌گی داده‌ها باعث می‌شود این نقطه به سختی قابل تشخیص باشد. یک نمونه بزرگتر حتماً می‌تواند مفیدتر باشد.

همه فرایندهای طبیعی (مانند همه آن‌هایی که توسط انسان ساخته شده) مشمول تغییر پذیری هستند. به عنوان مثال مهندسین عمران به این مطلب که مقاومت خرد شدن بتن، فشار خاک، مقاومت نقاط جوش، جریان ترافیک، سیلاب و میزان آلودگی در جریان‌های سطحی تغییرات گسترده‌ای دارند آگاه هستند. این ممکن است به دلیل تغییرات طبیعی در خواص، تفاوت در اثرات متقابل مابین عناصر تشکیل دهنده مواد، عوامل زیست محیطی یا علل دیگر باشد. برای مقابله با عدم قطعیت، مهندس در ابتدا باید نمونه‌ای از داده‌ها را گردآوری کرده و بررسی کند (مانند مجموعه‌ای از داده‌های جریان رودخانه و یا نتایج آزمایش سه محوری). نمونه در مرحله توصیفی مورد استفاده آمار و احتمال قرار می‌گیرد. با این حال برای اهداف استنباطی، نیاز به تصمیم‌گیری در مورد جامعه‌ای است که نمونه از آن بیرون کشیده شده است. منظور از جامعه کل یا مجموعی است که برای بسیاری از فرایندهای فیزیکی مجموعه تقریباً نامحدودی است که شامل تمام اندازه‌گیری‌های ممکن می‌شود. علاقه اصلی آمارگر به جمع اطلاعات است. اطلاعات تک عددی نکات، سرنخ‌ها و شواهد خوبی را فراهم می‌کنند.

مجموعه داده‌ها شامل تعدادی از اندازه‌گیری‌ها از یک پدیده مانند بار گسیختگی یک جزء ساختمانی است. مقادیر اندازه‌گیری شده، متغیر نامیده می‌شوند که هر یک ممکن است یک مقدار از مجموعه‌ای مشخص از مقادیر را اختیار کنند. به دلیل تصادفی بودن ذاتی و از آن رو غیر قابل پیش بینی بودن آن، پدیده‌ای که یک مهندس یا دانشمند معمولاً با آن سر و کار دارد به عنوان یک متغیر تصادفی خوانده می‌شود. این نام به هر کمیتی که مقدار آن به شانس بستگی دارد اطلاق می‌شود<sup>1</sup>. متغیرهای تصادفی معمولاً با حروف بزرگ نشان داده می‌شوند. طبقه‌بندی این متغیرها بر اساس نوع مقادیر ممکن (و یا فرض شده) برای آن‌ها است. الگوی تغییر پذیری متغیر توزیع نامیده می‌شود. متغیر پیوسته می‌تواند هر مقداری را در یک مقیاس پیوسته بین دو حد اختیار کند، مانند حجم جریان عبوری آب از رودخانه در هر ثانیه یا مقدار بارش روزانه اندازه‌گیری شده در برخی شهرها. در مقابل، یک متغیر گسسته تنها می‌تواند اعداد قابل شمارش مجزا مانند اعداد صحیح را اختیار کند مانند تعداد وسایل نقلیه‌ای که در یک تقاطع به سمت چپ می‌پیچند.

پس از به دست آمدن نمونه‌ای از داده‌ها، اولین گام نحوه ارائه آن است. برای مثال، جدول گسیختگی داده‌ها برای نوع خاصی از الوار در جدول 1.1 (ت) در پیوست نشان داده شده است. مشکل اولیه‌ای که مهندس با آن روبرو است این است که چنین مجموعه‌ای از داده‌ها به خودی خود هیچ ایده روشنی از ویژگی‌های اساسی مقدار تنش در این نوع از مصالح ساختمانی طبیعی به دست نمی‌دهد. برای استخراج ویژگی‌های برجسته و انواع خاصی از اطلاعات که مورد نیاز است، می‌بایست داده‌ها خلاصه شوند و در اشکالی که به آسانی قابل درک هستند ارائه گردند. روش‌های مختلفی برای ارائه، سازمان‌دهی و خلاصه‌کردن وجود دارد. روش‌های گرافیکی رویکرد اول هستند.

### 1.1 نمایش نموداری

اگر "یک تصویر ارزش هزار کلمه را دارد"، پس تکنیک‌های نموداری یک روش ایده‌آل برای تجسم تغییرپذیری و دیگر خواص مجموعه‌ای از داده‌ها است. با کمک سامانه قدرتمند تعاملی مغز و چشم، نمایش‌های نموداری درک و بینش کلی نسبت به فرم و شکل داده‌ها ارائه داده و منجر به ارائه یک مفهوم اولیه از فرآیند تولید داده‌ها می‌شود. با جمع‌آوری داده‌ها در قالب نمودار، می‌توان جزئیات داده‌ها را به طور اجمالی مرور کرد و ویژگی‌های مهم آن را برجسته ساخت. انواع متعددی از نمودار وجود دارد. نمودار خط و نقطه، هیستوگرام، چند ضلعی فراوانی نسبی و منحنی فراوانی تجمعی در این بخش آورده شده است. پس از آن، روش‌های کاوشگرانه مانند نمودار ساقه و برگ، نمودار جعبه‌ای و نمودارهایی که ارتباط احتمالی بین دو متغیر را به تصویر می‌کشند، در بخش 1.3 و 1.4 ارائه شده‌اند. ترسیم نمودار با کار ساده شمارش آغاز می‌شود.

<sup>1</sup> این اصطلاح به طور مفصل در بخش 3.1 تعریف شده است.



### 1.1.1 نمودار خطی یا نمودار میله‌ای

وقوع یک متغیر گسسته را می‌توان در قالب یک نمودار خطی و یا نمودار میله‌ای نشان داد. در این نوع از نمودار، محور افقی مقادیر متغیر گسسته و ارتفاع خطوط عمودی فراوانی را نشان می‌دهد. گسترده‌گی افقی این خطوط و ارتفاع نسبی آن تغییرپذیری و دیگر ویژگی‌های داده‌ها را نشان می‌دهد.

مثال 1.1. وقوع سیلاب. تعداد سالانه سیلاب رودخانه مگرا در کالامازا، بین پیزا و جنوا در شمال غرب ایتالیا طی یک دوره 34 ساله را همانطور که در جدول 1.1.1 نشان داده شده در نظر بگیرید. سیلاب در رودخانه در نقطه اندازه‌گیری به معنی بالا آمدن رودخانه بیش‌تر از یک سطح مشخص است که فراتر از آن رودخانه تهدیدی برای جان و مال مردم محسوب می‌شود. داده‌ها در شکل 1.1.1 به صورت یک نمودار خطی رسم شده‌اند. داده‌ها یک توزیع متقارن با مرکز چهار سیلاب در هر سال را نشان می‌دهند. در برخی دیگر حوزه‌ها فراوانی با افزایش تعداد سیلاب در سال (که از صفر شروع می‌شود) به صورت غیر خطی کاهش دارد که نشان دهنده یک مدل نمایی منفی تغییرات است.

### 1.1.2 نمودار نقطه‌ای

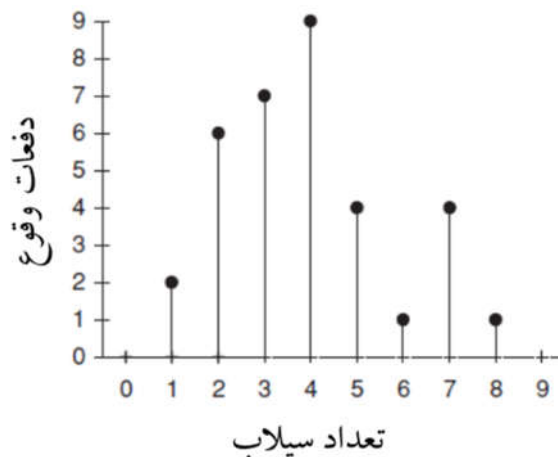
برای ارائه داده‌های پیوسته نوع متفاوتی از نمودار مورد نیاز است. اگر تعداد داده‌ها کم باشد (مثلاً کمتر از 25 مورد) نمودار نقطه‌ای یک کمک بصری بسیار مفید خواهد بود. تنها 15 مورد داده اول جدول 1.1.1 (ت) - که مدول گسیختگی را در واحد نیوتن بر میلی‌متر مربع برای چوب درخت ماموت سوئدی و چوب سفید در قطع  $50 \times 150$  میلی‌متر نشان می‌دهد - را در نظر بگیرید. این اطلاعات به صورت صعودی مرتب شده و در جدول 1.1.2 ارائه و در شکل 1.1.2 رسم شده است.

جدول 1.1.1 تعداد سالانه سیلاب رودخانه مگرا در کالامازا، بین پیزا و جنوا در شمال غرب ایتالیا طی دوره سال‌های 1939-1972<sup>a</sup>

تعداد وقوع	تعداد سالانه سیلاب
0	0
2	1
6	2
7	3
9	4
4	5
1	6
4	7
1	8
0	9
34	مجموع

«وقوع سیل به عنوان دبی بیش از 300 متر مکعب در ثانیه برای رودخانه تعریف شده است.»

می‌توان مشاهده کرد که مرکز داده‌ها نزدیک به 40 نیوتن بر میلی‌متر مربع است، اما گستردگی داده‌ها باعث می‌شود این نقطه به سختی قابل تشخیص باشد. یک نمونه بزرگتر حتماً می‌تواند مفیدتر باشد.



شکل 1.1.1 نمودار خطی برای وقوع سیلاب در رودخانه مگرا در کالامازا، بین پیزا و جنوا در شمال غرب ایتالیا.

### 1.1.3 هیستوگرام

اگر حداقل 25 مشاهده موجود باشد، یکی از رایج‌ترین اشکال نموداری یک نمودار بلوکی به نام هیستوگرام است. برای این منظور، داده‌ها با توجه به بزرگی‌اشان به چند گروه تقسیم می‌شوند. محور افقی نمودار نشان‌دهنده بزرگی مقادیر است. بلوک‌ها برای نشان دادن گروه‌ها ترسیم می‌شوند که هر یک از آن‌ها دارای یک حد بالا و پایین مشخص

هستند. مساحت هر بلوک متناسب با تعداد تکرار وقایع در گروه است. تغییرپذیری داده‌ها با پهنای افقی بلوک نشان داده شده است و مقادیر پرتکرار در بلوک با بیشترین مساحت یافت می‌شود. ویژگی‌های دیگری مانند تقارن داده‌ها و یا عدم آن نیز نشان داده می‌شود.

گام اول محاسبه دامنه تغییرات مشاهدات  $r$  است که برابر تفاوت بین بزرگترین و کوچکترین مقدار است. مثال 1.2. مقاومت الوار. به اطلاعات مقاومت الوار داده شده در جدول 1.1 (ت) برمی‌گردیم. این داده‌ها به ترتیب بزرگی در جدول 1.1.3 مرتب شده‌اند.

تعداد  $n = 165$  مشاهده با تغییرپذیری نسبتاً بالا وجود دارد (همانطور که انتظار می‌رود)، چرا که چوب ماده‌ای است که به طور طبیعی بسیار متغیر است. در این مورد دامنه تغییرات برابر  $r = 70.22 - 0.00 = 70.22$  نیوتن بر میلی‌متر مربع است.

برای رسم هیستوگرام، دامنه تغییرات داده‌ها به تعدادی رده یا سلول ( $nc$ ) تقسیم می‌کند. تعداد وقایع مشمول هر رده شمارش شده و در جدول یادداشت می‌شود. این تعداد فراوانی نامیده می‌شود.

پهنای رده‌ها معمولاً برای آسانی تفسیر یکسان در نظر گرفته می‌شوند. با این حال، برای برخی کاربردها از قبیل برازش یک تابع نظری به فراوانی‌های مشاهده شده، پهنای رده نابرابر استفاده می‌شود. در انتخاب تعداد رده‌ها ( $nc$ ) باید دقت شود. تعداد رده بیش از حد کم باعث حذف برخی از ویژگی‌های مهم داده‌ها می‌شود؛ تعداد رده بیش از حد زیاد تصویر کلی روشنی را از داده‌ها به دست نمی‌دهد چرا که ممکن است نوسانات زیادی در فراوانی‌ها به وجود آید.

جدول 1.1.2 پانزده داده اول مدول گسیختگی در آزمایش مقاومت الوار ارائه شده در جدول 1.1 (ت) (در واحد نیوتن بر میلی‌متر مربع). این داده‌ها با شروع از صفر به صورت صعودی مرتب شده‌اند.

29/11 29/93 32/02 32/40 33/06 34/12 35/58 39/34  
65/35 50/98 48/78 48/37 45/54 41/64 40/53



مدول گسیختگی بر حسب  $N/mm^2$

شکل 1.1.2 نمودار نقطه‌ای برای یک نمونه کوچک از مقاومت الوار از جدول 1.1.3.

یک قاعده سرانگشتی این است که تعداد رده‌ها برابر  $\sqrt{n}$  یا یک عدد صحیح نزدیک به آن باشد. از طرفی می‌بایست این تعداد حداقل 5 بوده و بیشتر از 25 نباشد. بنابراین، نمودار هیستوگرام که بر اساس کمتر از 25 داده ساخته شود ممکن است اصلاً معنی‌دار نباشد. استراگس (1926) تقریب زیر را پیشنهاد داد:

$$nc = 1 + 3/3 \log_{10} n. \quad (1.1.1)$$

فریدمن و دیاکونیس (1981) یک جایگزین برای این رابطه ارائه دادند که بیشتر بر اساس مبانی نظری به وجود آمده است:<sup>2</sup>

$$n_c = \frac{rn^{1/3}}{2iqr} \quad (1.1.2)$$

در اینجا  $iqr$  دامنه بین چارکی است. برای روشن شدن این مفهوم، می‌بایست  $Q_2$  یا همان میانه تعریف شود. میانه داده‌ای است که اگر مجموعه داده‌ها به ترتیب صعودی مرتب شوند در وسط قرار می‌گیرد و یا اگر  $n$  (تعداد داده‌ها) زوج باشد برابر میانگین دو داده وسطی است. چارک اول یا پایین  $Q_1$ ، میانه نیمه پایین داده‌ها است و به همین ترتیب چارک سوم یا بالا  $Q_3$ ، میانه نیمه بالایی داده‌ها است. از این پس این تعاریف استفاده می‌شوند.<sup>3</sup> بنابراین

$$iqr = Q_3 - Q_1. \quad (1.1.3)$$

<sup>2</sup> همچنین به اسکات (1979) رجوع کنید.

<sup>3</sup> جایگزین‌هایی مانند مقدار گرد شده  $(n+1)/4$  و  $(n+1) \times (3/4)$  به نزدیکترین عدد صحیح به ترتیب برای محاسبه مکان  $Q_1$  و  $Q_3$  وجود دارد. اگر عدد محاسبه شده از روابط بالا عددی صحیح نباشد، برای این دو رابطه گرد کردن به ترتیب به سمت بالا و پایین انجام می‌گیرد.

جدول 1.1.3 داده‌های مدول گسیختگی در آزمایش مقاومت الوار به واحد نیوتن در میلی‌متر مربع به ترتیب

صعودی<sup>a</sup>

0/00	28/00	31/60	34/44	36/84	39/21	41/75	44/30	47/25	53/99
17/98	28/13	32/02	34/49	36/85	39/33	41/78	44/36	47/42	54/04
22/67	28/46	32/03	34/56	36/88	39/34	41/85	44/36	47/61	54/71
22/74	28/69	32/40	34/63	36/92	39/60	42/31	44/51	47/74	55/23
22/75	28/71	32/48	35/03	37/51	39/62	42/47	44/54	47/83	56/60
23/14	28/76	32/68	35/17	37/65	39/77	43/07	44/59	48/37	56/80
23/16	28/83	32/76	35/30	37/69	39/93	43/12	44/78	48/39	57/99
23/19	28/97	33/06	35/43	37/78	39/97	43/26	44/78	48/78	58/34
24/09	28/98	33/14	35/58	38/00	40/20	43/33	45/19	49/57	65/35
24/25	29/11	33/18	35/67	38/05	40/27	43/33	45/54	49/59	65/61
24/84	29/90	33/19	35/88	38/16	40/39	43/41	45/92	49/65	69/07
25/39	29/93	33/47	35/89	38/64	40/53	43/48	45/97	50/91	70/22
25/98	30/02	33/61	36/00	38/71	40/71	43/48	46/01	50/98	
26/63	30/05	33/71	36/38	38/81	40/85	43/64	46/33	51/39	
27/31	30/33	33/92	36/47	<u>39/05</u>	40/85	43/99	46/50	51/90	
27/90	30/53	34/12	36/53	39/15	41/64	44/00	46/86	53/00	
27/93	31/33	34/40	36/81	39/20	41/72	44/07	46/99	53/63	

<sup>a</sup> مجموعه داده های اصلی در جدول 1.1 (پ) داده شده است؛  $n = 165$  . زیر عدد میانه خط کشیده شده است.  
مجموعه داده اصلی در جدول 1.1 (ث) داده شده است.

جدول 1.1.4 محاسبات فراوانی برای داده‌های مدول گسیختگی مرتب شده در جدول 1.1.3<sup>a</sup>

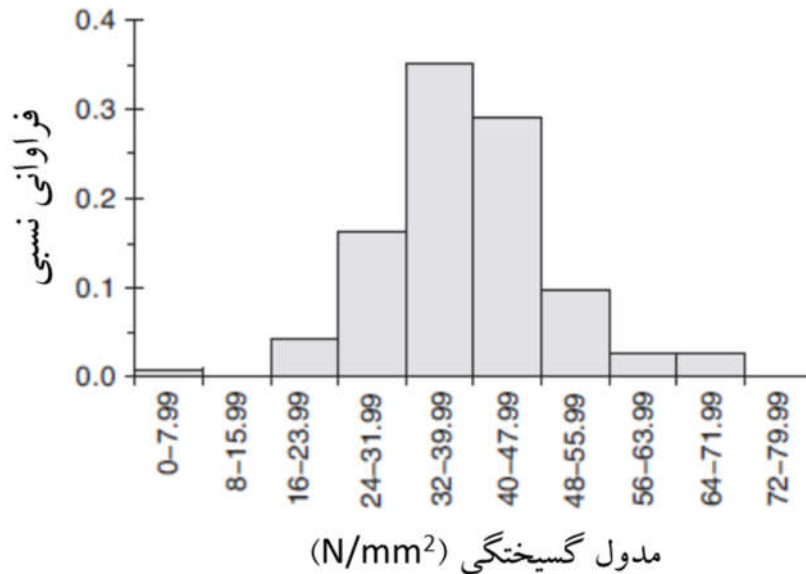
حد بالایی رده (N/mm <sup>2</sup> )	مرکز رده (N/mm <sup>2</sup> )	فراوانی مطلق	فراوانی نسبی	فراوانی نسبی تجمعی (%)
5	2/5	1	0/006	0/61
10	7/5	0	0	0/61
15	12/5	0	0	0/61
20	17/5	1	0/006	1/21
25	22/5	9	0/055	6/67
30	27/5	18	0/109	17/58
35	32/5	26	0/158	33/33
40	37/5	38	0/23	56/36
45	42/5	34	0/206	76/97
50	47/5	20	0/121	89/09
55	52/5	9	0/055	94/55
60	57/5	5	0/03	97/58
65	62/5	0	0	97/58
70	67/5	3	0/018	99/39
75	72/5	1	0/006	100

<sup>a</sup>پهنای هر رده در این مثال برابر 5 نیوتن بر میلیمتر مربع است.

مثال 1.3. **مقاومت الوار.** برای داده‌های مقاومت الوار جدول 1.1(ث)، میانه یا  $Q_2$  برابر 39/05 نیوتن بر میلیمتر مربع است. همچنین  $Q_3$  و  $Q_1$ ، به ترتیب برابر 44/57 و 32/91 نیوتن بر میلیمتر مربع بوده و از این رو  $iqr=11/66$  نیوتن بر میلیمتر مربع است. از قانون ساده جذر ارائه شده برای تعداد رده‌ها داریم،  $nc = 12/84$ . با این حال، با استفاده از روابط 1.1.1 و 1.1.2 تعداد رده‌ها برابر 8/32 و 16/52 به دست می‌آید. اگر این اعداد به 9 و 15 گرد شده و دامنه تغییرات برای اهداف نموداری به 72 و 75 نیوتن بر میلیمتر مربع گسترش یابد، پهنای رده برابر برای همه رده‌ها به ترتیب به 8 و 5 نیوتن بر میلیمتر مربع تبدیل می‌شود. فرض می‌شود که این مقادیر به عنوان پهنای رده استفاده شود. مهم است که مرز رده‌ها برای شمارش فراوانی بدون هیچگونه ابهامی مشخص شود؛ برای مثال، در مورد اول، رده باید از 0 تا 7/99، از 8/00 تا 15/99 و به همین ترتیب باشد. همانطور که قبلاً اشاره شد، محور عمودی هیستوگرام فراوانی و محور افقی به عنوان مقیاس اندازمگیری که در آن مرز رده‌ها علامت گذاری شده‌اند مورد استفاده قرار می‌گیرد. برای هر یک از پهنای رده 8 و 5 نیوتن بر میلیمتر مربع، حدود رده‌ها تعیین شده و شمارش فراوانی با استفاده از جدول 1.1.3 به انجام رسیده است؛ کوچکترین مرز 0 و بزرگترین آن به ترتیب در 72 و 75 نیوتن بر میلیمتر مربع است. جدول 1.1.4 فراوانی‌های مطلق و نسبی را برای پهنای رده 5 نیوتن بر میلیمتر مربع ارائه می‌دهد.

مستطیل‌هایی بر روی محدوده هر رده ساخته می‌شود که مساحت آن متناسب با فراوانی در رده است. هنگامی که همانند اینجا پهنای رده برابر استفاده شود، ارتفاع مستطیل بیانگر فراوانی است. بنابراین نمودارهای 1.1.3 و 1.1.4 به دست می‌آیند.

اطلاعات منتقل شده توسط دو نمودار هیستوگرام مشابه به نظر می‌رسند. نمودارها تقریباً متقارن است و اوج آن در رده زیر 40 نیوتن بر میلی‌متر مربع می‌باشد و کاهش مداومی در هر دو طرف اوج دارد. این نوع نمودار معمولاً هر نقص احتمالی در داده‌ها، مانند رخنه در اول یا آخر را برجسته می‌کند. تحقیقات بیشتری برای درک ماهیت واقعی جامعه آماری مورد نیاز است. در مورد این جنبه‌ها در همین فصل و فصل‌های بعدی بیشتر گفته خواهد شد.

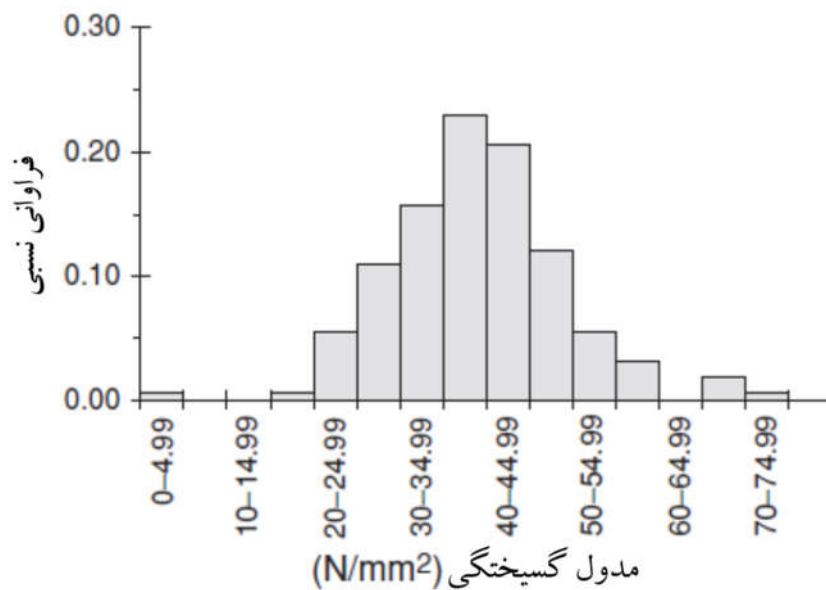


شکل 1.1.3 هیستوگرام داده‌های مقاومت الوار با پهنای رده 8 نیوتن بر میلی‌متر مربع.

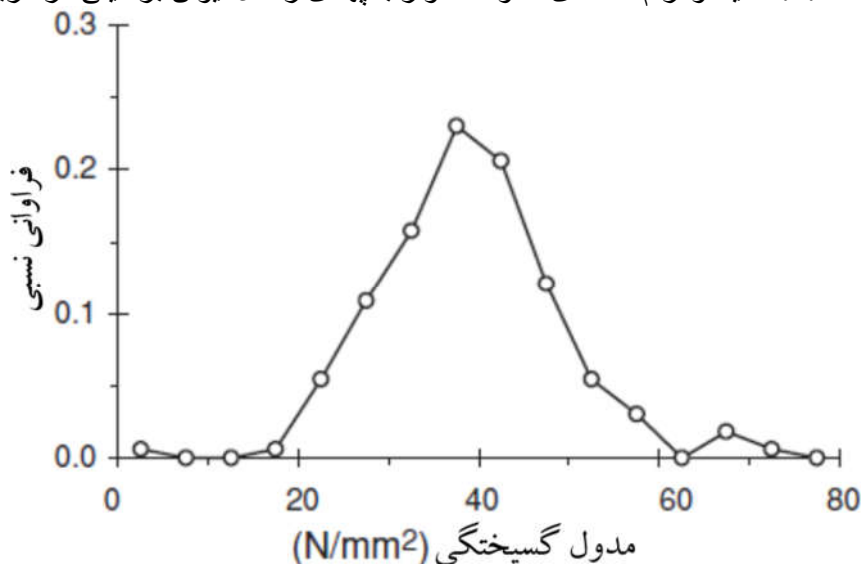
#### 1.1.4 چند ضلعی‌های فراوانی

چند ضلعی فراوانی یک ابزار تشخیصی مفید برای تعیین توزیع یک متغیر است. می‌توان پس از گسترش هیستوگرام به میزان یک رده در اول و یک رده در آخر، چند ضلعی [فراوانی] را با به هم پیوستن نقاط وسط بالای مستطیل‌های یک هیستوگرام به دست آورد. فرض می‌شود که پهنای رده برابر استفاده شود. در این صورت اگر ارتفاع هیستوگرام به تعداد کل مشاهدات تقسیم شود، هیستوگرام فراوانی نسبی به دست می‌آید. بنابراین، ارتفاع هر رده بیانگر احتمال آن است که بین 0 و 1 محدود شده است که همان شانس وقوع است. نمودار به نام چند ضلعی فراوانی نسبی خوانده می‌شود.

مثال 1.4. مقاومت الوار. هیستوگرام شکل 1.1.4 را در نظر بگیرید. مقادیر مرکز رده‌ها محاسبه شده و چند ضلعی فراوانی نسبی به دست آمده است که در شکل 1.1.5 نشان داده شده است.



شکل 1.1.4 هیستوگرام داده‌های مقاومت الوار با پهنای رده 5 نیوتن بر میلی‌متر مربع.



شکل 1.1.5 چند ضلعی فراوانی نسبی برای داده‌های مقاومت الوار با پهنای رده 5 نیوتن بر میلی‌متر مربع. همانطور که تعداد مشاهدات افزایش می‌یابد، پهنای رده از لحاظ نظری تمایل به کاهش دارد و در حالت حدی برای یک نمونه با اندازه بی‌نهایت، یک چند ضلعی فراوانی نسبی تبدیل به یک منحنی فراوانی می‌شود. در واقع این منحنی احتمال نشان دهنده یک تابع چگالی احتمال ریاضی است که به صورت مختصر به عنوان pdf جامعه خوانده می‌شود<sup>4</sup>.

#### 1.1.5 نمودار فراوانی نسبی تجمعی

اگر مقادیر فراوانی نسبی گام به گام از کوچکترین رده به بزرگترین رده جمع زده شود، خط حاصل از به هم

<sup>4</sup> این تابع در فصل 3 مورد بحث قرار گرفته است. یکی از اولین اهداف استفاده از آمار استنباطی، همانطور که در فصل‌های 4 و 5 معرفی خواهد شد، برآورد تابع ریاضی از یک نمونه محدود و بررسی نیکویی برازش آن به هیستوگرام مربوطه است.

پیوستن ارتفاع مرز بالایی رده‌ها که همان فراوانی‌های نسبی تجمعی هستند، نمودار فراوانی نسبی تجمعی و یا نمودار احتمال را تشکیل می‌دهد. این نمودار نشان دهنده احتمال عدم تجاوز از مقادیر مشخص شده بر روی محور افقی است. در عمل این نمودار با به کار گرفتن هر مورد از داده‌ها به صورت مجزا، بدون ضرورت ترسیم هیستوگرام و فرضیات محدود کننده‌ای که مشمول آن است رسم می‌شود. برای این منظور، می‌توان به سادگی تعداد مشاهدات کمتر یا مساوی هر یک از مقادیر را مشخص و این اعداد را به تعداد کل مشاهدات تقسیم نمود (به عنوان مثال، از داده‌های مرتب شده از جدول 1.1.3). این رویه در این جا نیز استفاده شده است.<sup>5</sup>

بنابر این نمودار احتمال که با نمودار فراوانی نسبی تجمعی نشان داده می‌شود، یک ابزار مهم عملی محسوب می‌شود. این نمودار میانه و دیگر چارک‌ها را به طور مستقیم ارائه می‌دهد. همچنین، می‌توان 9 مقداری که فراوانی کل را به 10 قسمت مساوی به نام دهک تقسیم می‌کنند معین کرد و یا مقادیر صدک را که در آن صدک  $p$ ام مقداری است که بزرگتر از  $p$  درصد از مشاهدات است تعیین نمود. به طور کلی، می‌توان  $(n-1)$  مقداری که کل فراوانی مشاهدات را به  $n$  قسمت مساوی تقسیم می‌کنند (=چندک) را پیدا کرد. از این رو یک چند ضلعی فراوانی تجمعی نیز یک نمودار چندکی یا نمودار-Q نامیده می‌شود، هر چند یک نمودار-Q بر خلاف نمودار فراوانی تجمعی روی محور عمودی خود مقادیر چندک را دارد.

مثال 1.5. مقاومت الوار. شکل 1.1.6 نمودار فراوانی تجمعی به دست آمده از داده‌های مرتب شده مقاومت الوار جدول 1.1.3 با استفاده از هر یک از داده‌ها (همانطور که در بالا شرح داده شد) است. دهک و صدک را می‌توان از آن استخراج کرد. بر طبق قرارداد احتمال یا یک مقیاس متناسب به جای درصد داده‌ها برای محور عمودی استفاده می‌شود (به جز در منحنی مدت که بعداً بحث خواهد شد). به عنوان مثال صدک 90، تقریباً برابر 51 نیوتن بر میلیمتر مربع و مقدار 40 نیوتن بر میلیمتر مربع احتمال عدم تجاویز برابر با حدود 0/56 دارد.

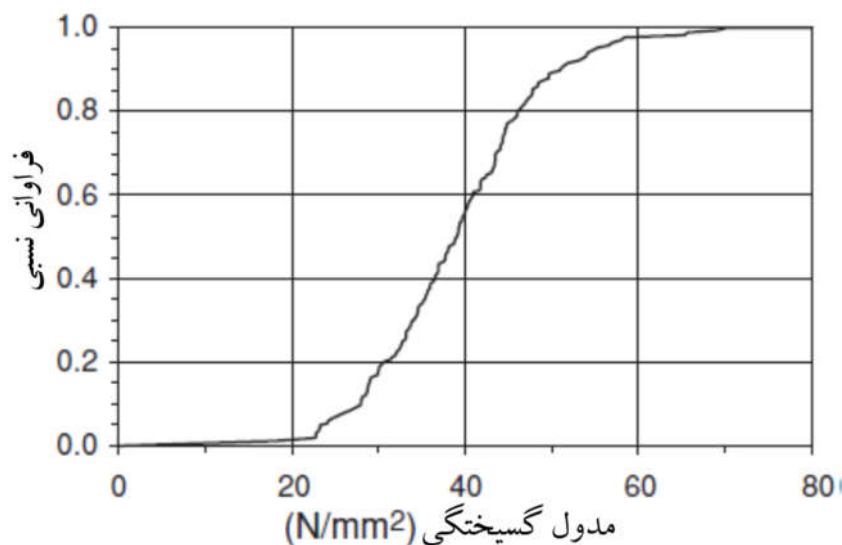
اگر اندازه نمونه به طور نامحدود افزایش یابد، در حالت حدی نمودار فراوانی نسبی تجمعی - دقیقاً به همان صورتی که یک چند ضلعی فراوانی نسبی منجر به یک تابع چگالی احتمال می‌شود - به منحنی توزیع تبدیل خواهد شد. این منحنی بیانگر جامعه آماری با استفاده از یک تابع توزیع (ریاضی) است که معمولاً به نام تابع توزیع تجمعی خوانده می‌شود (به صورت خلاصه شده cdf).

به عنوان یک روش گرافیکی برای مشخص کردن توزیع جامعه آماری، نمودار چندک را می‌توان با استفاده از یک مقیاس غیر خطی تغییر یافته برای مقادیر احتمال که نشان دهنده یکی از انواع مختلف توزیع‌های نظری است ترسیم کرد.<sup>6</sup> همچنین، همانطور که در بخش 1.4 نشان داده شده است دو توزیع را می‌توان با استفاده از یک نمودار Q-Q مقایسه کرد.

<sup>5</sup> جنبه‌های بیشتری از این موضوع که به نمودارهای احتمالی مربوط است در فصل 5 توضیح داده شده است.

<sup>6</sup> این روش در بخش 5.8 شرح داده شده است.



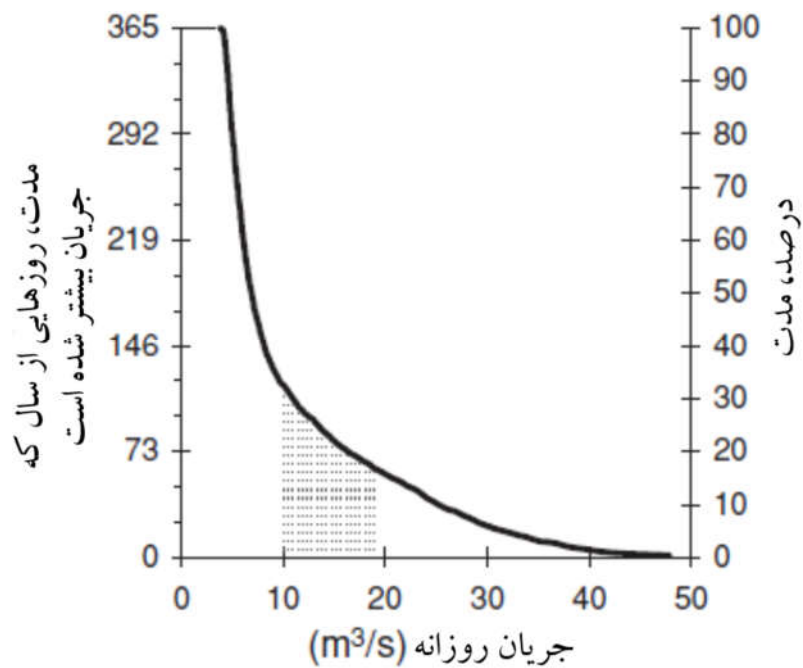


شکل 1.1.6 نمودار فرآوانی نسبی تجمعی برای داده‌های مقاومت الوار.

#### 1.1.6 منحنی مدت

برای ارزیابی منابع آب، طراحی‌های مربوط به آن و به منظور برنامه ریزی، رسم منحنی مدت زمان برای مهندسی بسیار مفید است. هنگامی که با جریان در رودخانه‌ها سر و کار داریم، این نوع از نمودار به عنوان منحنی مدت زمان جریان خوانده می‌شود. این منحنی در واقع ناشی از یک نمودار فرآوانی تجمعی با مقیاس‌های زمانی خاص است. به عنوان مثال محور عمودی می‌تواند نشان‌دهنده درصدی از زمان که جریان بیش از یک حد خاص است باشد و یا تعداد روز در سال یا فصل که در آن جریان بیشتر از یک حد (یا کمتر از آن) است ممکن است داده شود. حجم جریان در روز بر روی محور افقی داده شده است. برای برخی از اهداف، همانطور که در نمودار Q- معمول بود محورهای افقی و عمودی بعضاً جابه‌جا می‌شوند. یک مثال از کاربرد عملی این نمودار، مساحت منطقه محصور شده توسط یک منحنی است. یک خط افقی به نمایندگی از 100 درصد از زمان و یک خط عمودی در حداقل مقدار جریان که مطلوب است در رودخانه حفظ شود، کشیده شده است. این منطقه نشان‌دهنده حجم برآورد شده آب است که می‌بایست به صورت سالانه در رودخانه جاری باشد تا آن اهداف تامین شود.

مثال 1.6. مدت زمان جریان. شکل 1.1.7 منحنی مدت زمان رودخانه دورا ریپاریا را در منطقه کوهستانی آلپ در شمال ایتالیا که بر مبنای یک دوره 47 ساله از اطلاعات ایستگاه اندازه‌گیری سالبرترند ترسیم شده است نشان می‌دهد. این نمودار با استفاده از همان روش استفاده شده برای نمودار فرآوانی نسبی تجمعی، مانند شکل 1.1.6 رسم شده است. به عنوان مثال، فرض کنید که تصمیم گرفته شده است که برای دبی‌های بالاتر از 10 مترمکعب در ثانیه و زیر 20 مترمکعب در ثانیه قسمتی از جریان رودخانه منحرف و منتقل شود. مساحت ناحیه محدود شده توسط منحنی و خطوط عمودی کشیده شده در این دبی‌ها، با استفاده از مقیاس عمودی در سمت چپ، پس از ضرب شدن در تعداد ثانیه در روز حداکثر مقدار برآورد شده برای آب در دسترس برای انتقال در سال را برحسب مترمکعب به دست می‌دهد. این منطقه در شکل 1.1.7 سایه‌دار است. اگر چنین تصمیمی در یک دوره بلند مدت اجرا شود، می‌بایست از یک سری طولانی مدت داده‌ها استفاده شده و تابع توزیع برآورد شود.



شکل 1.1.7 منحنی مدت زمان جریان رودخانه دورا ریپاریا در سالبرترند در منطقه کوهستانی آلپ ایتالیا.