

گروه‌بندی و پردازش داده‌های آماری

- هدفهای رفتاری: در پایان این فصل، فراگیر باید بتواند:
- ۱- توضیح دهد که پردازش آماری به چه منظوری انجام می‌گیرد.
 - ۲- مفهوم گروه‌بندی و نقش آن را برای توصیف آماری توضیح دهد.
 - ۳- نتایج مشاهدات را در یک جامعه آماری گروه‌بندی نماید.
 - ۴- مفهوم فراوانی را توضیح دهد.
 - ۵- توزیع فراوانیهای صفت متغیر را در جامعه بیان کند.
 - ۶- جدول «توزیع فراوانیهای صفت متغیر» را با استفاده از نتایج مشاهدات، تشکیل دهد.
 - ۷- نمودار جدول «توزیع فراوانیهای صفت متغیر» را رسم کند.
 - ۸- مفهوم «فراوانی انباشته» را تعریف نماید.
 - ۹- براساس نتایج مشاهدات، فراوانیهای انباشته را محاسبه کرده، توزیع فراوانیهای انباشته را تشکیل دهد و نقش آن را به‌عنوان مشخصه صفت متغیر توضیح دهد.
 - ۱۰- لزوم واردکردن فاصله‌ها را برای مقادیر یک صفت متغیر نشان دهد.
 - ۱۱- لزوم واردکردن «مفهوم چگالی فراوانیها» را بیان کند.
 - ۱۲- جدول توزیع چگالی فراوانی را تشکیل داده، نمودار آن را نیز رسم کند.

پردازش داده‌های آماری - توزیع صفت متغیر

پردازش داده‌های آماری^۱

داده‌های آماری که در مرحله مشاهده به‌دست آمده است، اعداد یا ارقامی هستند که هیچ‌گونه

مفهوم خاصی نداشته و نیاز به پردازش دارند تا آنها را به صورت یک مجموعه ادغام شده درآورد. پردازش نتایج مشاهدات، شامل منظم کردن، طبقه بندی یا گروه بندی داده ها، تشکیل جداول، محاسبه سرجمعها و مشخصه های عددی است که آن را «مرحله پردازش» می نامند. بنابراین هدف از پردازش، ادغام نتایج مشاهدات، یا تراکم کردن اطلاعات، به دست آوردن مشخصه های کلی به منظور توصیف اجمالی خصوصیات جامعه ها می باشد.

توزیع صفت متغیر

مطالعه صفت در جامعه، بیشتر به منظور آشکار ساختن تغییر پذیری صفت متغیر صورت می گیرد. برای این کار لازم است عناصر جامعه در گروه های یکسان دسته بندی شوند. گروه بندی نتایج مشاهدات، یکی از اساسی ترین روشهای پردازش و تحلیل اولیه اطلاعات آماری است.

اگر صفت متغیر مورد بررسی کمی گسسته باشد، مقادیر مشاهده شده یکسان را که با یک عدد بیان می شوند به عنوان یک گروه در نظر گرفته، تعداد آنها را مشخص می کنیم. سپس تمامی مقادیر مختلف صفت متغیر را به ترتیب صعودی مرتب نموده در مقابل هریک از این مقادیر تعداد آنها (فراوانی) را می نویسیم. نتیجه گروه بندی به صورت زیر به دست می آید:

$$X: x_1 \quad x_2 \quad \dots \quad x_k$$

$$F_i: F_1 \quad F_2 \quad \dots \quad F_k$$

دو مجموعه مقادیر صفت متغیر X و فراوانیها را با هم «توزیع فراوانیهای صفت متغیر» یا به طور ساده «توزیع صفت متغیر» می نامند.

تعداد عناصر را در هر گروه، «فراوانی مطلق^۱» آن گروه می نامند و بر طبق معمول آن را با F_i نشان می دهند.

واضح است که برای فراوانیهای مطلق در یک توزیع صفت، همواره تساوی زیر برقرار است:

$$F_1 + F_2 + F_3 + \dots + F_k = N \quad (1)$$

به عبارت دیگر، مجموع فراوانیهای مطلق صفت متغیر در جامعه، همواره برابر N یعنی حجم جامعه می باشد.

۱- Frequency

۲- Absolute Frequency

توزیع صفت متغیر، چگونگی تغییرات فراوانیها را برحسب تغییرپذیری صفت متغیر منعکس می‌نماید. توزیع صفت متغیر را با جدول نیز به صورت زیر می‌توان بیان نمود:

جدول ۱

X	x_1	x_2	x_3	x_k	—
F_i	F_1	F_2	F_3	F_k	N

چنین جدولی را «جدول توزیع فراوانیهای مطلق» می‌نامند.

مثال ۱: فرض کنیم جامعه مورد مطالعه $N=20$ خانوار باشد که صفت متغیر X در آن تعداد اعضای هر خانوار است. برای هر خانوار، تعداد اعضای آن با پرسش تعیین شده و نتایج مشاهدات به صورت زیر به دست آمده است:

$X: 2, 4, 3, 4, 4, 5, 2, 2, 4, 5, 5, 3, 4, 4, 3, 3, 2, 4, 3, 5$

اگر مقادیر مختلف صفت متغیر X را که به ترتیب ۲، ۳، ۴ و ۵ می‌باشند، در ۴ گروه طبقه‌بندی نموده، تعداد تکرار آنها (فراوانیها) را در برابر آنها قرار دهیم، گروههایی به صورت زیر به دست می‌آیند (جدول ۲):

جدول ۲

X تعداد اعضای خانوار	۲	۳	۴	۵	—
F_i تعداد خانوار (فراوانی)	۴	۵	۷	۴	$F_i = 20$

اگر فراوانیهای مطلق (F_i) را برحجم جامعه (N) تقسیم کنیم، توزیع فراوانیهای نسبی^۱ را به دست خواهیم آورد. فراوانیهای نسبی را با نماد f_i نشان خواهیم داد که از رابطه زیر به دست می‌آید:

$$f_i = \frac{F_i}{N} \quad (2)$$

فراوانی نسبی گروه نام در جدول توزیع فراوانیها، سهم هریک از مقادیر صفت متغیر را در کل جامعه نشان می‌دهد.

۱- Relative Frequency

حاصل جمع فراوانیهای نسبی در جدول توزیع فراوانی، همواره برابر ۱ می‌باشد.

$$f_1 + f_2 + f_3 + \dots + f_k = \sum_{i=1}^k f_i = 1 \quad (3)$$

توزیع فراوانیهای نسبی صفت متغیر را با جدول نیز می‌توان نشان داد :

جدول ۳- توزیع فراوانیهای نسبی صفت متغیر X

X	x_1	x_2	x_3	x_k	-
f_i	f_1	f_2	f_3	f_k	$f_i = 1$

برای مثال ۱، توزیع فراوانیهای نسبی را محاسبه می‌کنیم (جدول ۴) :

جدول ۴

X تعداد اعضای خانوار	۲	۳	۴	۵	-
f_i نسبت خانوارها در جامعه	$\frac{4}{20}$	$\frac{5}{20}$	$\frac{7}{20}$	$\frac{4}{20}$	۱

مثال ۲: سن کارگران یک کارخانه نساجی مورد پرسش واقع شده، نتایج به‌دست آمده در جدول زیر، گروه‌بندی شده است. می‌خواهیم جدول توزیع فراوانی نسبی سن کارگران را به‌دست آوریم :

جدول ۵

X سن	۱۹	۲۰	۲۲	۲۵	۲۶	۳۱	۳۴	۳۷	-
F_i فراوانی مطلق	۷	۶	۹	۱۰	۱۵	۱۲	۸	۱۳	۸۰

کافی است هریک از فراوانیهای مطلق (F_i) را به N (در اینجا $N=80$) تقسیم نماییم تا فراوانی نسبی هر گروه از سن کارگران به‌دست آید. نتیجه محاسبات در جدول ۶ آورده شده است :

جدول ۶

X سن	۱۹	۲۰	۲۲	۲۵	۲۶	۳۱	۳۴	۳۷	-
f_i فراوانی نسبی	$\frac{7}{80}$	$\frac{6}{80}$	$\frac{9}{80}$	$\frac{10}{80}$	$\frac{15}{80}$	$\frac{12}{80}$	$\frac{8}{80}$	$\frac{13}{80}$	$\frac{80}{80} = 1$

ملاحظه می‌شود جمع فراوانیهای نسبی برابر ۱ خواهد شد.

مفهوم فراوانی نسبی بهتر از فراوانی مطلق می‌تواند وضعیت صفت را در جامعه مشخص سازد، برای مثال اگر بگوییم $\frac{1}{8}$ کارگران سنی برابر ۲۵ سال دارند، بهتر است از اینکه بگوییم در این کارخانه ۱۰ کارگر ۲۵ ساله وجود دارد. چون فراوانی مطلق معلوم نمی‌کند از چند کارگر موجود در این کارخانه، ۱۰ نفر ۲۵ سال سن دارند، ولی فراوانی نسبی دقیقاً آن را مشخص می‌کند.

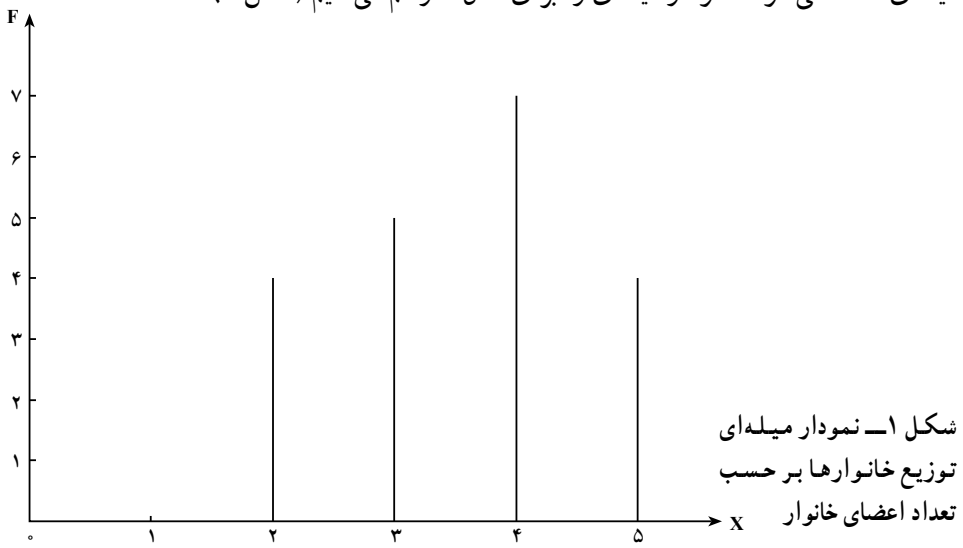
بیان هندسی توزیع صفت متغیر (نمودار میله‌ای^۱ – نمودار چندگوش^۲)

نمایش تغییرپذیری صفت متغیر با جداول توزیع فراوانیها و پی بردن به خصوصیات آن به‌ویژه وقتی این‌گونه جداول مفصل باشند، ساده نیست به طوری که از روی آنها به‌آسانی نمی‌توان تغییرات فراوانیها را از یک مقدار به مقدار دیگر صفت فهمید.

برای آشکار کردن آنها در آمار، غالباً از نمودار میله‌ای و نمودار چندگوش، استفاده می‌شود.

(البته این کار بیشتر اوقات زمانی که صفت متغیر کمی گسسته باشد صورت می‌گیرد).

برای ساختن نمودار میله‌ای در دستگاه مختصات (قائم)، بر روی محور طولها، مقادیر متغیر X و بر روی محور عرضها، فراوانیها، قرار داده می‌شوند. هر زوج F_i و X_i از توزیع فراوانیهای صفت متغیر، در دستگاه مختصات به صورت یک نقطه با طول X_i و عرض F_i نشان داده می‌شود. پس از رسم تمامی نقاط بر روی دستگاه مختصات، از نقاط به دست آمده، عمودهایی به محور X ها رسم می‌کنیم. به این ترتیب تصویری به دست می‌آید که از چند میله تشکیل شده است و به آن «نمودار میله‌ای» گفته می‌شود. نمودار میله‌ای را برای مثال ۱ رسم می‌کنیم (شکل ۱).

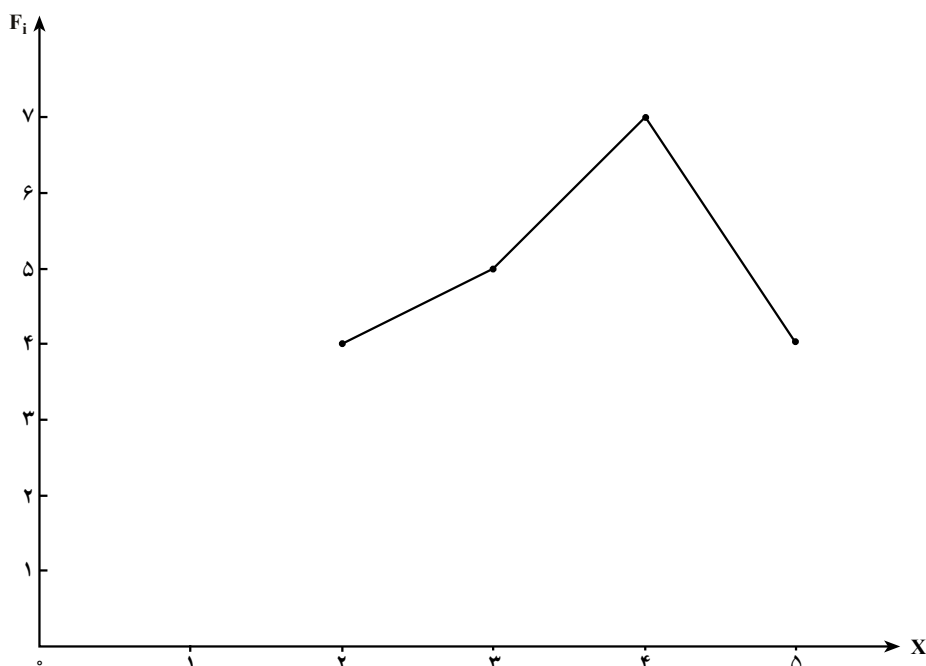


شکل ۱- نمودار میله‌ای
توزیع خانوارها بر حسب
تعداد اعضای خانوار

۱- Bar Diagram

۲- Polygon

با نمودار میله‌ای به سهولت می‌توان دریافت که خانوارهای ۴ نفره تعدادشان بیش از همه است و بعد از آن خانوارهای ۳ نفره و سپس خانوارهای ۲ و ۵ نفره به تعداد یکسان در این جامعه وجود دارند. برای رسم نمودار چندگوش، کافی است نقاط به دست آمده در دستگاه مختصات را با خطوط مستقیم به ترتیب به هم وصل نمود. در نتیجه شکلی حاصل می‌شود که از چند گوشه (زاویه) تشکیل شده است و به همین دلیل به آن «نمودار چندگوش توزیع فراوانیها» گفته می‌شود. این نوع نمودار بیشتر برای مقایسه توزیع فراوانیها در دو یا چند جامعه باهم مورد استفاده قرار می‌گیرد. چندگوش توزیع فراوانیها را برای مثال ۱ رسم می‌کنیم (شکل ۲):



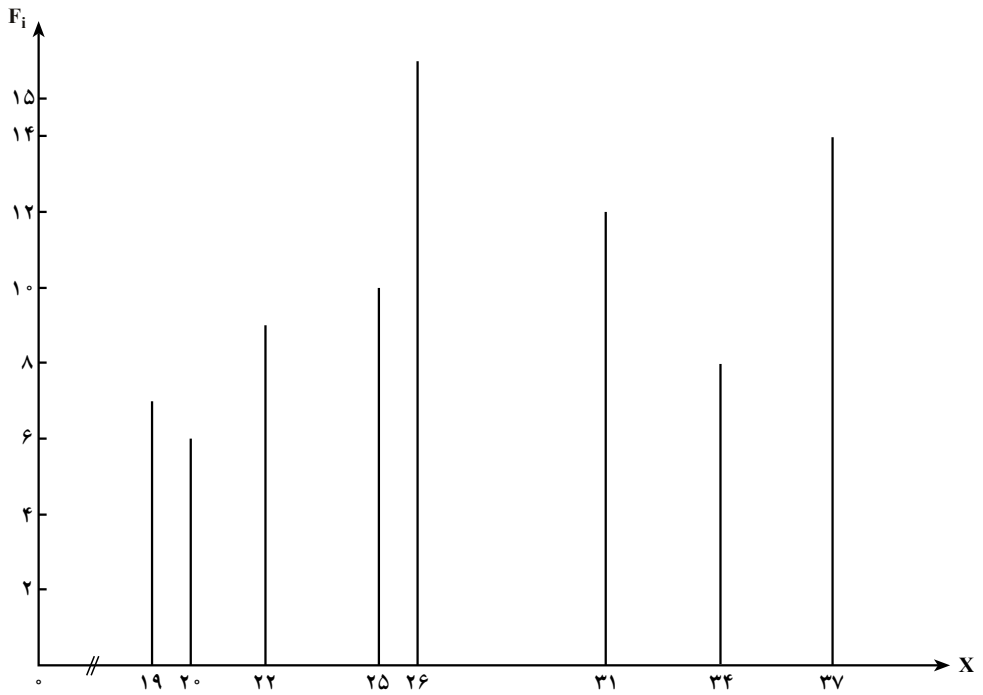
شکل ۲- نمودار چندگوش توزیع فراوانیهای تعداد اعضای خانوار

نمودار میله‌ای و چندگوش توزیع فراوانیها را برای مثال ۲ (سن کارگران کارخانه نساجی) رسم می‌کنیم (شکل‌های ۳ و ۴):

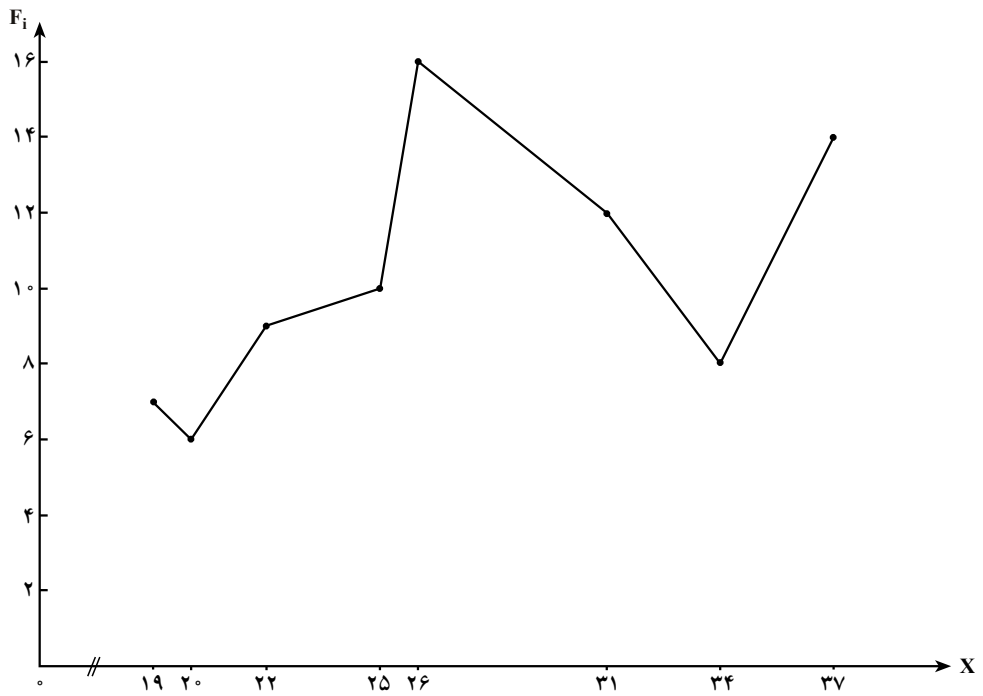
فراوانیهای تجمعی^۱ (انباشته)

فراوانیهای تجمعی نیز مانند فراوانیهای مطلق و نسبی، حجم گروهی از عناصر جامعه را بیان

^۱ - Cumulative Frequency



شکل ۳- نمودار میله‌ای سن کارگران کارخانه نساجی



شکل ۴- نمودار چندگوش سن کارگران کارخانه نساجی

می کند که به عنوان مشخصه صفت متغیر، مورد استفاده قرار می گیرد و به صورت زیر تعریف می شود :

فراوانی تجمعی هر گروه از صفت متغیر در جدول توزیع فراوانی، برابر است با فراوانی مطلق همان گروه به علاوه فراوانیهای مطلق گروههای ماقبل آن. فراوانی تجمعی (انباشته) مطلق را با نماد FC_i یا CF_i نشان می دهند.

اگر فراوانیهای مطلق را از ابتدا در جدول برای گروههای متوالی جمع کنیم، نهایتاً آخرین فراوانی انباشته (تجمعی) برابر با N یعنی حجم جامعه خواهد شد.
 فرم کلی محاسبات فراوانی انباشته به شکل زیر است (جدول ۷) :

جدول ۷

X_i	F_i	FC_i
x_1	F_1	$FC_1 = F_1$
x_2	F_2	$FC_2 = F_1 + F_2$
x_3	F_3	$FC_3 = F_1 + F_2 + F_3$
x_k	F_k	$FC_k = F_1 + F_2 + F_3 + \dots + F_k = N$
—	N	—

فراوانیهای انباشته را برای مثال ۱ محاسبه می کنیم (جدول ۸) :

جدول ۸

X_i	F_i	FC_i
۲	۴	$FC_1 = ۴$
۳	۵	$FC_2 = ۴ + ۵ = ۹$
۴	۷	$FC_3 = ۴ + ۵ + ۷ = ۱۶$
۵	۴	$FC_4 = ۴ + ۵ + ۷ + ۴ = ۲۰$
—	۲۰	—

ملاحظه می‌شود که آخرین فراوانی انباشته در جدول برابر ۲۰ یعنی حجم جامعه می‌باشد. همین مفهوم به صورت نسبت، به نام فراوانیهای انباشته نسبی نیز بیان می‌شود که به صورت زیر تعریف می‌گردد:

نسبت هریک از فراوانیهای تجمعی (انباشته) در جدول توزیع فراوانی را به N (حجم جامعه) فراوانی انباشته نسبی همان گروه صفت متغیر می‌نامند و آنرا با نماد fc_i یا fc_i نشان می‌دهند.

$$fc_i = \frac{FC_i}{N} \quad (4)$$

از حاصل جمع تجمعی فراوانیهای نسبی در جدول توزیع فراوانی، به همان ترتیبی که فراوانیهای انباشته مطلق را محاسبه نمودیم نیز می‌توان فراوانی انباشته نسبی هریک از گروههای صفت متغیر را به دست آورد. جدول توزیع فراوانیهای انباشته مطلق و نسبی را برای مثال ۲ محاسبه می‌کنیم:

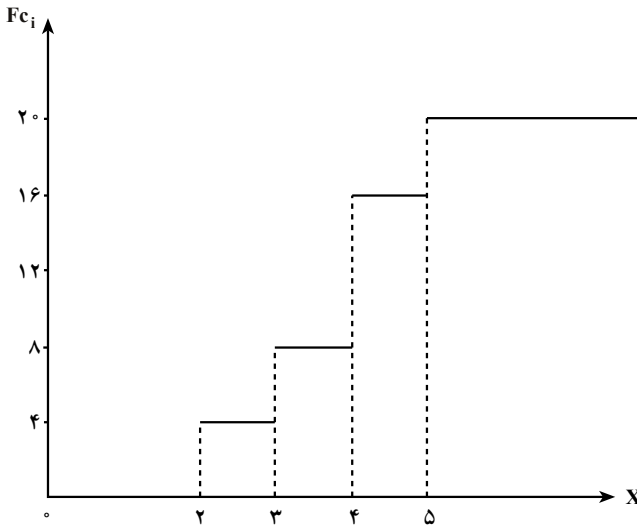
جدول ۹

X_i	F_i	f_i	FC_i	fc_i
۱۹	۷	$\frac{۷}{۸۰}$	۷	$\frac{۷}{۸۰}$
۲۰	۶	$\frac{۶}{۸۰}$	$۷+۶=۱۳$	$\frac{۷}{۸۰} + \frac{۶}{۸۰} = \frac{۱۳}{۸۰}$
۲۲	۹	$\frac{۹}{۸۰}$	$۱۳+۹=۲۲$	$\frac{۱۳}{۸۰} + \frac{۹}{۸۰} = \frac{۲۲}{۸۰}$
۲۵	۱۰	$\frac{۱۰}{۸۰}$	$۲۲+۱۰=۳۲$	$\frac{۲۲}{۸۰} + \frac{۱۰}{۸۰} = \frac{۳۲}{۸۰}$
۲۶	۱۵	$\frac{۱۵}{۸۰}$	$۳۲+۱۵=۴۷$	$\frac{۳۲}{۸۰} + \frac{۱۵}{۸۰} = \frac{۴۷}{۸۰}$
۳۱	۱۲	$\frac{۱۲}{۸۰}$	$۴۷+۱۲=۵۹$	$\frac{۴۷}{۸۰} + \frac{۱۲}{۸۰} = \frac{۵۹}{۸۰}$
۳۴	۸	$\frac{۸}{۸۰}$	$۵۹+۸=۶۷$	$\frac{۵۹}{۸۰} + \frac{۸}{۸۰} = \frac{۶۷}{۸۰}$
۳۷	۱۳	$\frac{۱۳}{۸۰}$	$۶۷+۱۳=۸۰$	$\frac{۶۷}{۸۰} + \frac{۱۳}{۸۰} = \frac{۸۰}{۸۰} = ۱$
—	۸۰	۱	—	$\frac{۸۰}{۸۰} = ۱$

فراوانی انباشته برای بیان تعداد مقادیر صفت x معین که از یک مقدار مشخص تجاوز نکنند، به کار می‌رود. در مثال ۲ اگر بخواهیم بدانیم مجموع کارگرانی که سن آنها از ۲۶ سال تجاوز نمی‌کند،

چه تعدادی است، کافی است فراوانی انباشته ردیف ۲۶ سالگی را مشاهده کرد و آن را به دست آورد، که در اینجا این رقم برابر ۴۷ می باشد. همچنین، اگر بخواهیم نسبت کارگرانی را که سن آنها از ۲۶ سالگی تجاوز نمی کند، بدانیم، کافی است، فراوانی انباشته نسبی سن ۲۶ سالگی را در جدول قرائت نماییم. این نسبت در جدول برابر $\frac{47}{80}$ می باشد.

توزیع فراوانیهای انباشته را نیز می توان به صورت نمودار بیان نمود. برای این کار روی محور طولها، مقادیر X و روی محور عرضها، فراوانیهای انباشته مطلق یا نسبی را قرار می دهیم و برای هر زوج X_i و FC_i یا fc_i یک نقطه در صفحه مختصات به دست می آوریم و سپس با وصل کردن این نقاط به شکلی دست می یابیم که به آن «نمودار توزیع فراوانیهای تجمعی» (کومولات^۱) گویند. نمودار فراوانیهای انباشته را برای مثال ۱ رسم می کنیم (شکل ۵):



شکل ۵- نمودار فراوانیهای تجمعی تعداد اعضای خانوار

نمودار فراوانیهای انباشته را برای مثال ۲ رسم می کنیم (شکل ۶):

طریقه عملی گروه بندی نتایج مشاهدات (داده های آماری)

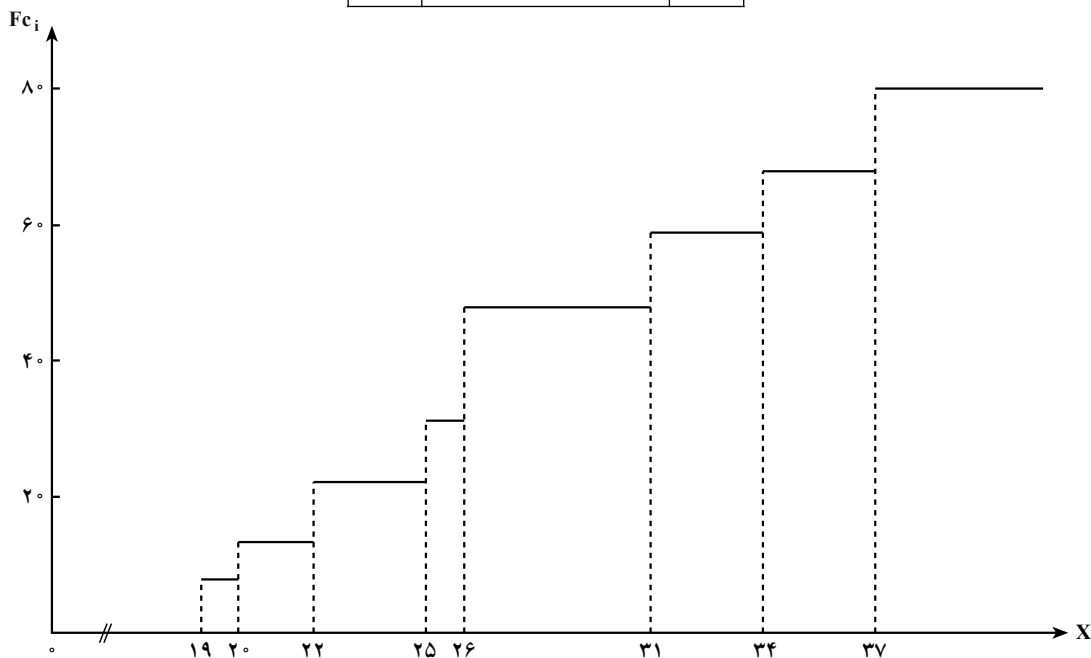
در مطالعه جامعه های با حجم کوچک، پردازش داده ها به صورت دستی انجام می گیرد. ولی در مطالعه جامعه های با حجم بزرگ، از پردازش ماشینی یا رایانه استفاده می شود. ما در اینجا به طور خلاصه فقط درباره پردازش دستی در مرحله گروه بندی بحث می کنیم.

۱- Cumulat

— روش چوب خط: در روش چوب خط، جدولها به صورت زیر تشکیل می گردد:

جدول ۱۰

X	ستون چوب خط	F_i
x_1	//	F_1
x_2	///	F_2
x_k	/	F_k
—	—	N



شکل ۶- نمودار فراوانیهای انباشته سن کارگران کارخانه نساجی

در ستون اول جدول، مقادیر صفت متغیر را به طور صعودی از کوچک به بزرگ قرار می دهیم. در ستون دوم، برای هر یک از مشاهدات در گروه X مربوط، یک چوب خط رسم می کنیم. این کار تا انتقال همه مشاهدات به صورت چوب خط ادامه می یابد. سپس تعداد چوب خطها را در هر گروه، شمارش کرده، نتیجه را به صورت فراوانی آن گروه در ستون سوم قرار می دهیم. در نهایت جمع فراوانیها برابر کل مشاهدات یعنی N خواهد شد.

اگر بخواهیم جدول فراوانیهای فاصله‌ای^۱ داشته باشیم، نحوه عملیات، مشابه جدول ۱۰ خواهد بود با این تفاوت که در ستون اول، مقادیر صفت متغیر را به صورت فاصله‌ها قرار می‌دهیم و سپس تک تک مشاهدات را در برابر فاصله مربوط با چوب خط رسم نموده، از شمارش آنها، فراوانیهای مطلق هر گروه را به دست خواهیم آورد که در ستون سوم نوشته می‌شود. به این ترتیب، جدول توزیع فراوانیهای فاصله‌ای برای صفت متغیر به دست می‌آید (جدول ۱۱):

جدول ۱۱

فاصله‌های صفت متغیر	ستون چوب خط	فراوانیها
$X_1 - X_2$	///	F_1
$X_2 - X_3$	////-	F_2
$X_3 - X_4$	//////	F_3
$X_k - X_{k+1}$	//	F_k
-	-	N

— مراحل تشکیل جدول توزیع فراوانیهای فاصله‌ای: مراحل تنظیم جدول توزیع فراوانیهای فاصله‌ای را به صورت زیر بیان می‌کنیم:

۱- پس از انجام مرحله مشاهده بزرگترین اندازه صفت متغیر (X_{max}) و کوچکترین اندازه آن را (X_{min}) مشخص کرده، تفاضل آنها را محاسبه می‌کنیم. یعنی

$$X_{max} - X_{min} = R \quad (5)$$

این تفاضل را که دامنه تغییرات^۲ نامیده می‌شود با R نشان می‌دهند.

۲- تعیین تعداد طبقات جدول - با توجه به حجم داده‌ها (N) تعداد طبقات را به دلخواه بین ۵ تا ۲۵ طبقه در نظر می‌گیرند. یکی از روشهای کلاسیک برای تعیین تعداد طبقات استفاده از فرمول استروجنس است که به صورت زیر می‌باشد:

۱- Interval Frequency

۲- Range = دامنه

$$K = 1 + 3 / 32 \log N \quad (6)$$

۳- از تقسیم دامنه تغییرات بر تعداد طبقات، فاصله طبقات به دست می آید که آن را با نماد I نشان خواهیم داد.

$$I = \frac{R}{K} \quad (7)$$

۴- تعیین فراوانیهای مطلق - در این مرحله با کمک روش چوب خط، تعداد اعضای را که اندازه آنها مربوط به هر طبقه جدول است، مشخص می کنیم.

با یک مثال، هم طریقه ساختن جدول توزیع فراوانیهای فاصله ای و هم طریقه عملی گروه بندی نتایج مشاهدات را نشان می دهیم:

مثال ۳: تعداد کارگران یک مؤسسه تولیدی $N = 100$ نفر می باشند. به منظور مطالعه تغییر پذیری سن آنها، تمامی کارگران را مورد مشاهده قرار داده ایم. نتایج مشاهدات برای سن آنها، به قرار زیر به دست آمده است:

۱۸	۲۱	۲۱	۱۸	۴۷	۵۳	۳۹	۲۸	۲۳	۵۰
۳۲	۳۳	۳۶	۴۴	۳۳	۲۲	۲۵	۲۲	۱۸	۱۹
۳۷	۲۴	۳۵	۳۵	۱۸	۳۹	۳۷	۳۲	۳۰	۴۴
۳۵	۳۵	۲۴	۲۷	۲۶	۳۲	۳۶	۳۵	۲۸	۳۲
۳۰	۴۶	۴۸	۴۴	۴۹	۵۰	۱۹	۲۰	۱۹	۴۹
۴۷	۴۳	۳۰	۳۸	۳۷	۲۰	۱۸	۳۰	۳۹	۴۵
۴۴	۲۴	۵۳	۲۸	۵۰	۲۸	۲۸	۵۲	۲۶	۱۸
۲۶	۲۷	۳۶	۳۸	۳۸	۴۰	۴۱	۴۱	۴۸	۴۳
۴۳	۳۲	۳۳	۳۲	۴۷	۲۵	۴۳	۳۲	۳۰	۳۵
۴۷	۳۴	۳۶	۳۸	۴۰	۴۱	۲۹	۴۲	۵۰	۵۴

اگر نتایج مشاهدات را بر حسب مقادیر مختلف سن طبقه بندی نمایم، ۳۶ گروه سنی تشکیل خواهد شد، که به علت زیادی گروههای صفت X ، امکان نتیجه گیری نسبت به تغییر پذیری سن مشکل است. از این رو، در اینجا مناسب است از روش گروه بندی فاصله ای استفاده کنیم، یعنی با ادغام کردن

چند گروه سنی باهم، فاصله‌های سنی بسازیم. با این کار، تعداد گروه‌های صفت متغیر کمتر و از طولانی شدن جدول کاسته می‌شود.

برای این منظور، ابتدا دامنه تغییرات صفت را محاسبه می‌کنیم.

در اینجا بزرگترین اندازه صفت (سن) ۵۴ سالگی و کوچکترین اندازه آن، ۱۸ سالگی است،

بنابراین داریم :

$$54 - 18 = 36$$

دامنه تغییرات سن

اگر بخواهیم نتایج مشاهدات را در ۸ گروه طبقه‌بندی کنیم ($k=8$)، طول فاصله خواهد شد :

$$\frac{36}{8} = 4.5 \approx 5$$

طول فاصله

(در چنین مواردی بهتر است طول فاصله را به عدد صحیح تبدیل کنیم).

بنابراین طول فاصله‌هایی را که در جدول توزیع فراوانی تشکیل می‌دهیم، برابر ۵ سال در نظر

می‌گیریم. برای تشکیل اولین طبقه در جدول توزیع فراوانی، کوچکترین اندازه صفت را (X_{\min})

برابر کرانه پایین طبقه اول در نظر گرفته، با توجه به مقدار I (فاصله طبقات)، سایر طبقات جدول را

تنظیم می‌کنیم. توضیح اینکه کوچکترین اندازه هر طبقه را، کرانه پایین آن طبقه و بزرگترین اندازه آن

طبقه را، کرانه بالای آن طبقه نامند.

نتایج این گروه‌بندی در جدول زیر نشان داده شده است (جدول ۱۲) :

جدول ۱۲

طبقات سنی X	چوب خط	فراوانی F_i
۱۸-۲۲	//// //	۱۵
۲۳-۲۷	//// //	۱۱
۲۸-۳۲	//// //	۱۸
۳۳-۳۷	//// //	۱۷
۳۸-۴۲	//// //	۱۳
۴۳-۴۷	//// //	۱۴
۴۸-۵۲	////	۹
۵۳-۵۷	///	۳
-	-	۱۰۰

باید توجه داشت که هریک از فاصله‌ها از ۵ سال تشکیل شده‌اند. چون سن، یک صفت ناپیوسته است، بنابراین کرانه بالای هر گروه با کرانه پایین گروه بعدی یک واحد تفاوت دارد، به عبارت دیگر، سن ۲۲ سالگی در گروه اول قرار می‌گیرد و سن ۲۳ سالگی که بلافاصله بعد از ۲۲ سالگی قرار دارد، در گروه دوم منظور می‌شود. به این ترتیب، با قرائت تک‌تک اعداد که در مرحله مشاهده به دست آمده است و انتقال آنها به فاصله مربوط در جدول، به وسیله چوب خط، کلیه اعداد را در ستون دوم وارد می‌کنیم و از شمارش آنها، فراوانیهای هر گروه (هر فاصله) را به دست می‌آوریم که حاصل جمع آنها نیز برابر N یعنی ۱۰۰ خواهد بود.

اهمیت فراوانیهای فاصله‌ای و مناسب بودن آن، زمانی آشکار می‌گردد که صفت متغیر مورد مطالعه «پیوسته» باشد.

در مطالعه صفت متغیر پیوسته، فاصله‌ها را باید چنان انتخاب کرد که صفت در هریک از آنها تقریباً به طور یکنواخت توزیع شده باشد. معمولاً انجام این کار وقتی امکان‌پذیر است که فاصله‌های نامساوی انتخاب شوند. لیکن انتخاب فاصله‌های نامساوی، انجام عملیات بعدی را مشکل‌تر می‌سازد. از این رو، در عمل اکثر اوقات، نتایج مشاهدات را به وسیله توزیع فراوانیهای با فاصله‌های یکسان بیان می‌کنند.

در استفاده از فاصله‌های با طولهای متفاوت، تحلیل توزیع فراوانیهای فاصله‌ای مشکل‌تر می‌گردد، زیرا فراوانیهای گروهها نه تنها از خصوصیات صفت متغیر تبعیت می‌کند، بلکه از طول فاصله انتخاب شده نیز تأثیر خواهد پذیرفت.

در اینجا به این نتیجه می‌رسیم که در مطالعه صفت متغیر که به صورت توزیع فاصله‌ای بیان شده باشد، باید از مفهوم جدیدی به نام «چگالی فراوانی»^۱ استفاده کنیم.

چگالی فراوانی، فراوانی را در واحد اندازه صفت نشان می‌دهد.

یعنی:

$$dF_i = \frac{\text{فراوانی گروه } i}{\text{فاصله همان گروه}} = \frac{F_i}{\Delta x_i} \quad (8)$$

$$(9) \quad df_i = \frac{\text{فراوانی نسبی گروه } i}{\text{فاصله همان گروه}} = \frac{f_i}{\Delta x_i}$$

مثال ۴: فرض کنیم نتایج مشاهدات برای صفت متغیر، موجودی حسابهای پس انداز در یکی از شعب بانک باشد که برحسب فاصله‌ها بیان شده است (جدول ۱۳):

جدول ۱۳

(موجودی برحسب هزار تومان) X	۷-۱۰	۱۰-۱۵	۱۵-۲۵	۲۵-۴۰	۴۰-۶۰	-
F _i فراوانیها	۳۰	۴۰	۶۰	۶۰	۱۰	۲۰۰

اگرچه در توزیع فوق با فاصله ۷-۱۰، (به طول ۳ واحد)، فراوانی ۳۰ و با فاصله ۱۵-۱۰ (به طول ۵ واحد) فراوانی ۴۰ در تناظر است، اما به این معنا نیست که با تغییر صفت از فاصله ۷-۱۰ به فاصله ۱۵-۱۰ فراوانیها افزایش یافته، بلکه برعکس در فاصله اول به سهم هر واحد طول فاصله $\frac{۳۰}{۳} = ۱۰$ عنصر جامعه تعلق می‌گیرد، در صورتی که در فاصله دوم، به هر واحد طول فاصله فقط $\frac{۴۰}{۵} = ۸$ عنصر جامعه تعلق می‌گیرد.

چگالی فراوانیها را برای فراوانیهای فاصله‌ای محاسبه کرده، توزیع چگالی فراوانیها را به دست می‌آوریم.

جدول ۱۴

فاصله‌های صفت	۷-۱۰	۱۰-۱۵	۱۵-۲۵	۲۵-۴۰	۴۰-۶۰
چگالی فراوانی $df_i = \frac{F_i}{\Delta x_i}$	$\frac{۳۰}{۳} = ۱۰$	$\frac{۴۰}{۵} = ۸$	$\frac{۶۰}{۱۰} = ۶$	$\frac{۶۰}{۱۵} = ۴$	$\frac{۱۰}{۲۰} = ۰/۵$

ملاحظه می‌شود که چگالی فراوانیها با تغییر صفت متغیر از فاصله‌ای به فاصله دیگر تغییر کرده، نزول می‌یابد. این نزولی بودن تصادفی نیست. برای اطمینان، کافی است فاصله‌ها را عوض کنیم، باز هم خواهیم دید که تغییرات چگالی فراوانیها برای این صفت متغیر، نزولی است. اگر در توزیع چگالی فراوانیها، فاصله‌ها را تغییر دهیم، شکل کلی توزیع تغییر نخواهد کرد. بدین جهت، در مطالعه توزیع فراوانیهای فاصله‌ای، با فاصله‌های نامساوی، مناسب است، این توزیع فراوانیها، به توزیع چگالی فراوانیها تبدیل گردد.

مثال ۵: فرض کنیم نتایج مشاهدات برای یک صفت متغیر کمی پیوسته X ، به وسیله جدول توزیع فراوانیهای نسبی فاصله‌ای بیان شده باشد:

جدول ۱۵

فاصله‌های صفت X	۱۱-۱۵	۱۵-۲۵	۲۵-۲۸	۲۸-۳۵	-
فراوانیهای نسبی f_i	۰/۰۸	۰/۴۰	۰/۲۴	۰/۲۸	$f_i = 1$

ظاهراً در این توزیع، بین تغییرات فراوانیها با فاصله‌های صفت متغیر، نظم خاصی دیده نمی‌شود، و بزرگترین فراوانی نسبی متعلق به فاصله ۱۵-۲۵ است. حال اگر توزیع فراوانیها را به صورت توزیع چگالی فراوانیها، درآوریم، نتیجه به شکل جدول زیر به دست می‌آید:

جدول ۱۶

فاصله‌های صفت X	۱۱-۱۵	۱۵-۲۵	۲۵-۲۸	۲۸-۳۵	-
فراوانیهای نسبی f_i	۰/۰۸	۰/۴۰	۰/۲۴	۰/۲۸	$f_i = 1$
چگالی فراوانیهای نسبی df_i	۰/۰۲	۰/۰۴	۰/۰۸	۰/۰۴	-

از توزیع چگالی فراوانیها نتیجه می‌گیریم که حداکثر آن در فاصله ۱۵-۲۵ نیست، بلکه به فاصله ۲۵-۲۸ تعلق دارد. یعنی اکثر عناصر جامعه در این فاصله گروه‌بندی شده‌اند. بنابراین، هرگاه جدول توزیع فراوانی فاصله‌ای، با فاصله‌های نامساوی داشته باشیم، برای درک این مطلب که تراکم فراوانیها در چه فاصله‌ای از صفت متغیر قرار دارند، باید، ابتدا چگالی فراوانیهای صفت متغیر را محاسبه نموده، با استفاده از آنها، نسبت به تراکم فراوانیها نتیجه‌گیری کنیم.

بیان هندسی توزیع فراوانیهای فاصله‌ای (صفت متغیر پیوسته)

نمودار مستطیلی یا هیستوگرام^۱ (بافت‌نگار): برای بیان هندسی توزیع صفت پیوسته که با فاصله

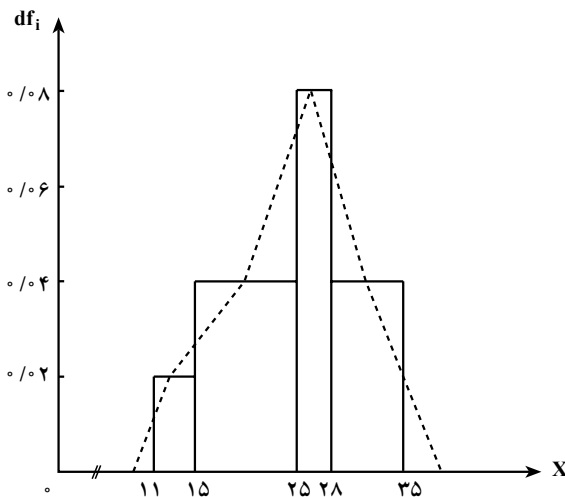
بیان شده است، از دو نوع نمودار استفاده می‌شود:

– نمودار مستطیلی یا هیستوگرام برای توزیع چگالی فراوانیها

– نمودار تجمعی یا اجایو برای توزیع فراوانیهای انباشته

در ساختن هر دو نوع نمودار، فرض بر این است که چگالی (تراکم) در هر یک از فاصله‌ها ثابت باشد.

برای ساختن هیستوگرام، بر روی محور طولها (در دستگاه مختصات قائم) فاصله‌های صفت و بر روی محور عرضها، چگالی فراوانیها را قرار می‌دهند. به عبارت دیگر بر روی فاصله‌های صفت متغیر، مستطیلهایی به ارتفاع مساوی با چگالی متناظر با آن فاصله‌ها می‌سازند. برای جدول ۱۶ در مثال ۵، هیستوگرام توزیع صفت متغیر X را رسم می‌کنیم.



شکل ۷- هیستوگرام جدول ۱۶

برای بیان هندسی توزیع صفتی که با فاصله‌ها بیان شده است، می‌توان از نمودار چندگوش نیز استفاده کرد. برای رسم چندگوش توزیع صفت، کافی است روی شکل هیستوگرام، نقاط وسط اضلاع فوقانی مستطیلهای را با خطوط مستقیم به هم وصل کنیم. از اتصال تمام نقاط وسط مستطیلهای به هم، نمودار چندگوش به دست می‌آید که مساحت زیر آن با محور X ها، برابر با مجموع مساحت مستطیلهای، می‌باشد.

در شکل ۷ چندگوش توزیع صفت به وسیله خط چین رسم شده است.

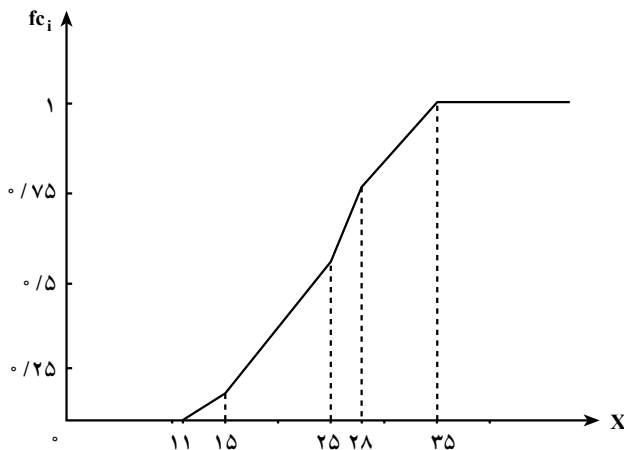
در ساختن نمودار تجمعی یا اجابو با فراوانیهای انباشته، از همان فرض ثابت بودن چگالی در فاصله‌های صفت استفاده می‌شود. در صفحه مختصات قائم، روی محور طولها، کرانه‌های هر فاصله را قرار داده، روی محور عرضها، فراوانیهای انباشته را درجه بندی می‌کنیم. به همان ترتیبی که برای صفت کمی گسسته، چندگوش فراوانیهای انباشته رسم گردید، عمل می‌کنیم و از به دست آوردن نقاط

(باطول X_i و عرض FC_i متناظر)، آنها را به وسیله خطوط مستقیم، به هم وصل می‌کنیم. در این صورت، شکلی ظاهر می‌شود که نمودار فراوانیهای انباشته نام دارد.

در زیر، جدول توزیع فراوانیهای انباشته نسبی را برای مثال ۵ تشکیل می‌دهیم و با استفاده از آن نمودار تجمعی توزیع صفت را رسم می‌کنیم (جدول ۱۷ و شکل ۸):

جدول ۱۷

X	۱۱-۱۵	۱۵-۲۵	۲۵-۲۸	۲۸-۳۵	-
f_i	۰/۰۸	۰/۴	۰/۲۴	۰/۲۸	۱
fc_i	۰/۰۸	۰/۴۸	۰/۷۲	۱	-



شکل ۸ - نمودار تجمعی (جدول ۱۷) توزیع صفت متغیر X

می‌توان نشان داد که مساحت هیستوگرام (جمع مساحت مستطیلهای) و همین‌طور مساحت زیرخط شکسته چندگوش، وقتی که برای رسم آنها از چگالی فراوانیهای مطلق استفاده شده باشد، همواره مساوی با N است و وقتی که برای رسم آنها از چگالی فراوانیهای نسبی استفاده شده باشد، مساحت هیستوگرام و همین‌طور مساحت زیرخط شکسته چندگوش، مساوی با ۱ است.

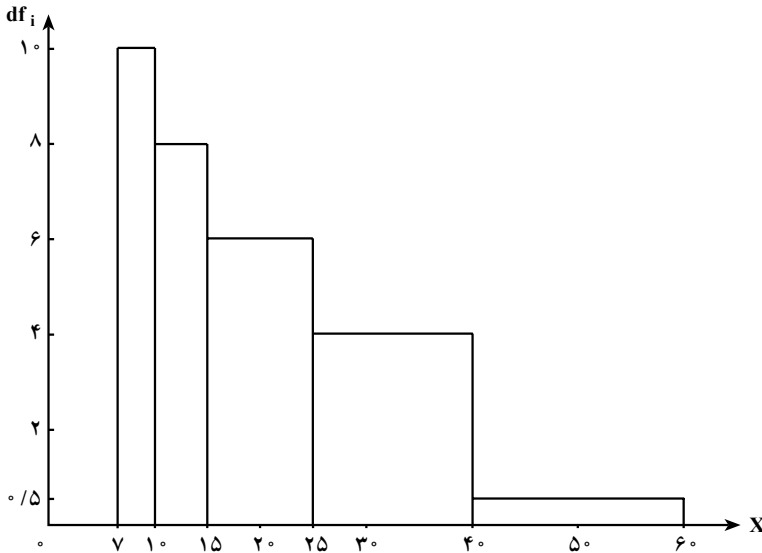
نمودارهای هیستوگرام و تجمعی توزیع صفت را برای مثال ۴ رسم می‌کنیم تا دانش‌آموزان با چگونگی رسم آنها آشنایی بیشتری پیدا کنند.

ابتدا، چگالی فراوانیهای مطلق و فراوانیهای انباشته توزیع صفت را محاسبه می‌کنیم.

جدول ۱۸

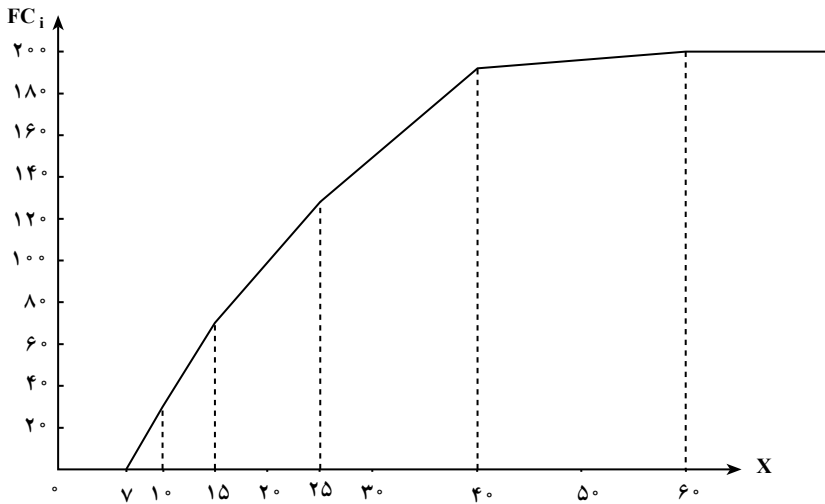
X	۷-۱۰	۱۰-۱۵	۱۵-۲۵	۲۵-۴۰	۴۰-۶۰	-
F_i	۳۰	۴۰	۶۰	۶۰	۱۰	۲۰۰
df_i	۱۰	۸	۶	۴	۰/۵	-
FC_i	۳۰	۷۰	۱۳۰	۱۹۰	۲۰۰	-

برای رسم هیستوگرام، ابتدا، فاصله‌های صفت را روی محور طولها، در دستگاه مختصات قائم وارد کرده، روی محور عرضها، چگالی فراوانیها را قرار می‌دهیم، سپس روی هر فاصله صفت متغیر، مستطیلی با طول چگالی فراوانی همان فاصله رسم می‌کنیم.



شکل ۹ - هیستوگرام توزیع صفت متغیر X (جدول ۱۸)

برای رسم نمودار تجمعی، اندازه‌های صفت را روی محور طولها و فراوانیهای انباشته را روی محور عرضها در دستگاه مختصات قرار می‌دهیم، سپس با هر مقدار صفت و فراوانی انباشته متناظر آن، یک نقطه روی صفحه مختصات به دست می‌آوریم. با وصل نمودن نقاط به دست آمده، به وسیله خطوط مستقیم، شکل نمودار تجمعی توزیع صفت متغیر به دست می‌آید.



شکل ۱۰- نمودار تجمعی توزیع صفت متغیر X (مربوط به مثال ۴)

سؤالها و تمرینها ؟

- ۱- لزوم گروه‌بندی نتایج مشاهدات در چیست؟
- ۲- فراوانیهای گروه، چه مفهومی دارد؟
- ۳- توزیع صفت متغیر، چیست؟
- ۴- فراوانی مطلق، چه مفهومی دارد؟
- ۵- فراوانی نسبی، چه مفهومی دارد؟
- ۶- بین فراوانیهای مطلق چه رابطه‌ای برقرار است؟
- ۷- بین فراوانیهای نسبی چه رابطه‌ای وجود دارد؟
- ۸- توزیع فراوانیهای صفت متغیر چیست؟ تعریف کنید.
- ۹- فرض کنیم صفت متغیر کمی ناپیوسته در جامعه فقط مقادیر ۵، ۱۰، ۱۵، ۲۰ و ۲۵ را به ترتیب با فراوانیهای ۵، ۱۰۰، ۲۰۰، ۱۰۰ و ۵۰ اختیار کرده باشد.
اولاً: توزیع فراوانیهای صفت متغیر فوق را بیان کنید.
ثانیاً: جدول توزیع فراوانیهای فوق را تشکیل دهید.
- ۱۰- توزیع فراوانیهای نسبی صفت متغیر ناپیوسته، چگونه به دست می‌آید؟
- ۱۱- برای صفت متغیر در تمرین ۹، فراوانیهای نسبی را محاسبه کنید.

- ۱۲- بین فراوانیهای مطلق و فراوانیهای نسبی چه رابطه‌ای برقرار است؟
- ۱۳- دلیل استفاده از نمودارها برای بیان توزیع صفت متغیر چیست؟
- ۱۴- نمودار میله‌ای، چگونه ساخته می‌شود؟
- ۱۵- برای توزیع صفت متغیر در تمرین ۹، نمودار میله‌ای رسم کنید.
- ۱۶- چندگوش توزیع فراوانیها چیست و چگونه ساخته می‌شود؟
- ۱۷- چندگوش توزیع فراوانیها را برای توزیع صفت متغیر که در جدول زیر داده شده است، رسم کنید.

X	۲	۴	۸	۱۰	۱۲	۱۴	-
F _i	۱۲	۲۴	۳۶	۴۸	۲۶	۴	۱۵۰

۱۸- توزیع صفت متغیر X با جدول زیر بیان شده است :

X	۳	۶	۹	۱۲	۱۵	-
f _i	۰/۱	۰/۳	۰/۳	۰/۲	۰/۱	۱

نمودار میله‌ای و نمودار چندگوش را برای توزیع فوق رسم کنید.

۱۹- فراوانی انباشته (تجمعی) چیست؟ تعریف کنید.

۲۰- توزیع فراوانیهای انباشته چیست؟

۲۱- توزیع فراوانی صفت متغیر، با جدول زیر بیان شده است :

X	۲	۴	۶	۸	۱۰	۱۲	-
F _i	۲۰	۴۰	۷۰	۳۵	۲۰	۱۵	۲۰۰

جدول توزیع فراوانیهای انباشته را برای صفت متغیر تشکیل دهید.

۲۲- فراوانی انباشته نسبی چیست و چگونه محاسبه می‌شود؟

۲۳- جدول توزیع فراوانیهای انباشته نسبی را برای تمرین ۲۱ تشکیل دهید.

۲۴- توزیع فراوانیهای انباشته به صورت هندسی چگونه بیان می‌شود؟

۲۵- نمودار تجمعی یا اجایو، برای صفت متغیر چگونه رسم می‌شود؟

۲۶- نمودار تجمعی را برای توزیع صفت در تمرین ۲۱ رسم کنید.

۲۷- چرا در گروه بندی نتایج مشاهدات صفت متغیر، از فاصله مقادیر صفت استفاده می شود؟

۲۸- لزوم تشکیل چنین فاصله هایی در چیست؟

۲۹- توزیع فراوانیهای فاصله ای، چیست؟

۳۰- وزن دانش آموزان کلاس چهارم دبستان A به شرح زیر اندازه گیری شده است.

۲۸/۵	۳۲/۵	۳۱/۷	۳۱/۸	۲۷/۹
۳۱/۴	۳۰/۸	۳۲/۵	۲۹/۶	۳۱/۷
۳۰/۶	۳۰/۴	۲۹/۳	۲۸/۷	۲۸/۹
۲۹/۳	۲۹/۷	۲۸/۸	۲۸/۶	۳۰/۵
۳۲/۸	۳۳/۲	۳۳/۵	۳۱/۴	۳۱/۶
۳۰/۲	۲۷/۸	۲۸/۲	۲۹/۳	۲۸/۲
۲۹/۸	۳۰/۸	۳۱/۲	۳۲/۸	۳۳/۴
۳۰/۲	۲۸/۶	۲۹/۹	۲۹/۳	۳۱/۸
۳۴/۱	۳۴/۳	۳۲/۴	۳۱/۲	۳۴/۴
۲۸/۷	۲۹/۴	۳۱/۴	۲۷/۸	۲۸/۲

الف - جدول توزیع فراوانیهای گروه بندی شده وزن دانش آموزان را در ۸ گروه تشکیل دهید.

ب - جدول توزیع فراوانیهای تجمعی وزن دانش آموزان را تشکیل دهید.

ج - چند درصد دانش آموزان وزنی بیش از ۳۰ کیلوگرم دارند؟

د - هیستوگرام وزن دانش آموزان را رسم کنید.

۳۱- در چه مواردی انتخاب فاصله های با طول نامساوی را به فاصله های با طول یکسان

ترجیح می دهید؟

۳۲- چگالی فراوانیها چیست و لزوم وارد کردن آن در چیست؟

۳۳- توزیع فراوانیهای کارگران بر حسب دستمزد روزانه در یک کارگاه ساختمانی به صورت

جدول زیر بیان شده است :

X دستمزد روزانه (صد تومان)	۱۶. ۲۵	۲۵. ۳۰	۳۰. ۳۵	۳۵. ۴۰	۴۰. ۵۰	.
F _i فراوانی	۲۲	۹۸	۶۰	۱۸	۲	۲۰۰

توزیع چگالی فراوانیها را برای دستمزد روزانه کارگران تشکیل دهید.

۳۴- دلیل ثبات چگالی فراوانیها در چیست؟

۳۵- هیستوگرام یا نمودار مستطیلی چیست؟ در چه مواردی از هیستوگرام استفاده می شود؟

- ۳۶- فرضی که در ساختن هیستوگرام باید در نظر گرفت، چیست؟
 ۳۷- برای تمرین ۳۳ هیستوگرام را رسم کنید.
 ۳۸- برای تمرین ۳۳ نمودار تجمعی یا اجایو را رسم کنید.

خودآزمونهای چهارگزینه‌ای ✓

- ۱- کدام رابطه بین فراوانیهای مطلق صحیح است؟
 الف - $F_i = 1$.
 ب - $F_i = \frac{N}{4}$.
 ج - $F_i = 0$.
 د - $F_i = N$.
- ۲- کدام رابطه بین فراوانیهای نسبی صحیح است؟
 الف - $f_i = 0$.
 ب - $f_i \neq 0/5$.
 ج - $f_i = N$.
 د - $f_i = 1$.
- ۳- فراوانی انباشته هر گروه از صفت متغیر، در جدول توزیع فراوانی برابر است با:
 الف - فراوانی مطلق گروه مربوط
 ب - فراوانی نسبی گروه مربوط
 ج - فراوانی گروه مربوط به علاوه فراوانیهای ماقبل آن
 د - N
- ۴- برای بیان هندسی صفت متغیر گسسته از کدام نمودارها، استفاده می‌شود؟
 الف - نمودار فراوانی تجمعی
 ب - نمودار میله‌ای و چندگوش
 ج - هیستوگرام و نمودار مستطیلی
 د - نمودار نقطه‌ای
- ۵- برای رسم هیستوگرام روی محورهای مختصات کدام مؤلفه‌ها قرار داده می‌شود؟
 الف - فاصله‌های صفت متغیر و چگالی فراوانی
 ب - اندازه صفت و فراوانی نسبی
 ج - فاصله‌های صفت و فراوانیهای انباشته
 د - فاصله‌های صفت و فراوانیهای مطلق
- ۶- چگالی فراوانی صفت متغیر یعنی:
 الف - فراوانی هر طبقه تقسیم بر N
 ب - فراوانی هر طبقه تقسیم بر فراوانی نسبی همان طبقه
 ج - فراوانی هر طبقه تقسیم بر فاصله آن طبقه
 د - فراوانی انباشته تقسیم بر N
- ۷- اگر هیستوگرام توزیع صفت با استفاده از چگالی فراوانیهای مطلق، رسم گردد، مساحت مستطیل‌های هیستوگرام برابر خواهد شد با:
 الف - ۱
 ب - $0/5$
 ج - N
 د - $\frac{N}{4}$