مدرسان شریف

فصل اول

آمار توصیفی

«تعاریف اولیه»

جمعیت یا جامعهٔ آماری: مجموعه تمام افراد یا اشیایی که مطالعات آماری در مورد یک یا چند صفت آنها در یک مکان و زمان معین انجام می‌گیرد، جامعه یا جمعیت آماری گفته می‌شود. تعداد اعضای جمعیت به حجم جامعه معروف است که آن را با N نمایش می‌دهند.

نمونه: در بررسیهای آماری به دلیل هزینه زیاد، کمبود وقت و در بعضی مواقع غیر ممکن بودن انجام کار زیر مجموعه‌ای از جمعیت، با قاعده و ضابطه خاصی انتخاب می‌شود که به آن نمونه گویند و بررسی بر روی آن نمونه انجام می‌گیرد. تعداد اعضای نمونه به حجم نمونه معروف است که آن را با n نمایش می‌دهند.

صفت مشخصه: صفتی است که بین همه‌ی عناصر جامعه‌ی آماری مشترک و جدا کننده جامعه آماری از سایر جوامع است. به طور کلی صفات خود، به دو دسته تقسیم می‌شوند: ۱ـ صفات کمی ۲ـ صفات کیفی

داده‌های آماری: در بررسی آماری، باید صفت مورد مطالعه به صورت اعداد و ارقام نمایش داده شود اگر صفت مورد مطالعه کمی باشد این عمل به سادگی امکان‌پذیر است. اما اگر صفت مورد مطالعه کیفی (مانند گروه خون، رنگ چشم و ...) باشد آنگاه باید طبق یک قاعده مشخص این صفات کیفی با عدد و رقم نمایش داده شوند. به طور کلی داده‌ها به دو صورت گسسته و پیوسته تقسیم می‌شوند.

داده‌های گسسته: داده‌هایی هستند که بین دو مقدار متصور آنها هیچ عددی وجود ندارد مانند تعداد فرزندان یک خانواده که مقادیر ۰,۱,۲,... است.

داده‌های پیوسته: داده‌هایی هستتند که بین هر دو مقدار متصور آنها همواره عدد دیگری وجود دارد مانند: وزن افراد ـ طول قد افراد ـ طول عمر قطعات الکتریکی و ... .

✍ **تست ۱:** ***کدامیک از صفات زیر یک صفت کمی است؟***

۱) گروههای خونی متفاوت ۲) رنگ چشم افراد ۳) عمر یک انسان ۴) هیچکدام

☑ **پاسخ: گزینه «۳»** گزینه (۱) و (۲) صفات کیفی هستند ولی عمر انسان یک صفت کمی است.

چگونه اعداد را گرد می‌کنیم

فرض کنیم عددی را می‌خواهیم تا n رقم اعشار گرد کنیم. تقریبی را که بر اساس آن می‌خواهیم عدد را گرد کنیم در نظر مـی‌گیـریم آن را نصف کرده و به عدد اصلی اضافه می‌کنیم سپس n رقم اعشار را نگه داشته و بقیه ارقام را صفر می‌کنیم.

برای درک بهتر به مثالهای زیر توجه کنید.

مثال ۱: ***عدد ۵۸۲۴/۰۱۵۷ را با تقریب ۰/۱ گرد کنید.***

☑ **پاسخ:** تقریب ۰/۱ را نصف می‌کنیم این مقدار به عدد اصلی اضافه یک رقم پشت ممیز را نگه داشته و بقیه را صفر می‌کنیم.

$$۵۸۲۴/۰۱۵۷ + ۰/۰۵ = ۵۸۲۴/۰۶۵۷ \rightarrow ۵۸۲۴/۰$$

مثال ۲: ***عدد ۷۵۲۳/۰۱۸۲ را با تقریب ۱ گرد کنید.***

☑ **پاسخ:** نصف عدد ۱ برابر با ۰/۵ است. پشت ممیز هیچ عددی را نگه نمی‌داریم.

$$۷۵۲۳/۰۱۸۲ + ۰/۵ = ۷۵۲۳/۵۱۸۲ \rightarrow ۷۵۲۳$$

مثال ۳: ***عدد ۶۷۲۳/۰۳۸۶ را با تقریب ۱۰۰ گرد کنید.***

☑ **پاسخ:** نصف عدد ۱۰۰ برابر با ۵۰ است.

$$۶۷۲۳/۰۳۸۶ + ۵۰ = ۶۷۷۳/۰۳۸۶ \rightarrow ۶۷۰۰$$

مثال ۴: ***عدد ۳۷۷۵/۰۱۵۲ را با تقریب ۰/۰۰۱ گرد کنید.***

☑ **پاسخ:** نصف عدد ۰/۰۰۱ برابر با ۰/۰۰۰۵ است.

$$۳۷۷۵/۰۱۵۲ + ۰/۰۰۰۵ = ۳۷۷۵/۰۱۵۷ \rightarrow ۳۷۷۵/۰۱۵$$

مفهوم و کاربرد نماد Σ

نماد Σ برای جمع کردن بکار برده می‌شود فرض کنید بخواهیم اعداد ۱ و ۲ و ۳ و ۴ و ۵ را با هم جمع کنیم در اینصورت می‌توانیم به جای استفاده از علامت + از نماد Σ که سیگما نامیده می‌شود استفاده کنیم.

$$\sum_{i=۱}^{۵} i = ۱+۲+۳+۴+۵$$

$$\sum_{i=۱}^{۸} i = ۱+۲+۳+۴+۵+۶+۷+۸$$

$$\sum_{i=۱}^{n} x_i = x_۱ + x_۲ + x_۳ + \ldots + x_n$$

خواص Σ

c یک عدد ثابت است.

۱) $\sum_{i=1}^{n} c = nc$

۲) $\sum_{i=1}^{n} cx_i = c\sum_{i=1}^{n} x_i$

۳) $\sum_{i=1}^{n} (x_i + c) = \sum_{i=1}^{n} x_i + nc$

۴) $\sum_{i=1}^{n} (ax_i + b) = a\sum_{i=1}^{n} x_i + nb$

۵) $\sum_{i=1}^{n} (ax_i + by_i) = a\sum_{i=1}^{n} x_i + b\sum_{i=1}^{n} y_i$

مثال ۵: ***فرض کنید متغیر X مقادیری مانند ۳ و ۵ و ۷ را بگیرد مقدار $\sum_{i=1}^{3} (x_i + 5)^2$ را بدست آورید.***

☑ **پاسخ :**

$$\sum_{i=1}^{3} (x_i + 5)^2 = (x_1 + 5)^2 + (x_2 + 5)^2 + (x_3 + 5)^2 = (3+5)^2 + (5+5)^2 + (7+5)^2 = 64 + 100 + 144 = 308$$

در آمار توصیفی بعد از جمع‌آوری داده‌ها مراحل زیر انجام می‌گیرد.

۱ـ تنظیم و طبقه بندی داده‌ها در یک جدول

۲ـ ترسیم نمودارهای گوناگون با استفاده از مقادیر جدول

۳ـ خلاصه کردن داده‌ها به یک یا چند عدد موسوم به شاخص یا آماره

در زیر این مراحل را توضیح خواهیم داد.

جدول فراوانی: متداولترین جدول آماری به جدول فراوانی معروف است. یک جدول فراوانی شامل موارد زیر است:

فراوانی و فراوانی نسبی: اگر n داده از k نوع داشته باشیم و تعداد این داده‌ها در این k طبقه به ترتیب $f_1, f_2, \ldots, f_k$ باشند به $f_1, f_2, \ldots, f_k$ فراوانی‌های مطلق طبقات و به $r_1 = \frac{f_1}{n}, r_2 = \frac{f_2}{n}, \ldots, r_k = \frac{f_k}{n}$ فراوانی‌های نسبی طبقات گویند.

$$1 \le f_i \le n \Rightarrow f_1 + f_2 + \ldots + f_k = \sum_{i=1}^{k} f_i = n$$

$$0 \le r_i \le 1 \Rightarrow r_1 + r_2 + \ldots + r_k = \sum_{i=1}^{k} r_i = 1$$

فراوانی تجمعی و فراوانی تجمعی نسبی: اگر در هر طبقه فراوانی آن طبقه و طبقات قبل از آن را با یکدیگر جمع کنیم فراوانی تجمعی آن طبقه بدست می‌آید و اگر در هر طبقه فراوانی نسبی آن طبقه و طبقات قبل را با هم جمع کنیم فراوانی تجمعی نسبی آن طبقه بدست می‌آید. پس:

فراوانی تجمعی طبقه jام $= g_j = f_1 + f_2 + \ldots + f_j$

فراوانی تجمعی نسبی طبقه jام $= s_j = r_1 + r_2 + \ldots + r_j$

مراحل ساخت یک جدول فراوانی برای داده‌های پیوسته

۱ـ دریافت داده‌های خام و در صورت لزوم گرد کردن آنها

۲ـ تقسیم داده‌ها به تعدادی رده یا طبقه. یک قاعده مفید می‌گوید: (n تعداد کل داده‌ها و k تعداد طبقات است) $k = ۱ + ۳/۳۲۲\log_{۱۰}^{n}$

۳ـ $\text{میزان تغییرپذیری داده‌ها} = s = \dfrac{\text{واحد گرد شده داده‌ها}}{۲}$

۴ـ دامنه داده‌ها که برابر با تفاضل کوچکترین داده از بزرگترین داده می‌باشد را بدست می‌آوریم. (R) $R = \max - \min$

۵ـ طول هر رده یا طبقه از تقسیم دامنه بر تعداد رده (k) بدست می‌آید. $w = \dfrac{R}{k}$

مثال ۶: *برای داده‌های زیر جدول فراوانی بسازید. (داده‌های زیر طول عمر نوعی قطعه الکتریکی برحسب ساعت می‌باشند.)*

۳۷ ـ۳۲ ـ۳۸ ـ ۳۲ ـ ۳۴ ـ ۳۲ ـ ۴۰ـ ۳۸ـ ۳۲ ـ ۳۵ ـ ۳۴ ـ ۲۹ ـ ۳۶ـ ۳۷ـ ۲۴ـ ۳۵ ـ ۳۹ ـ ۳۳ـ ۴۲ـ ۴۵ ـ ۳۰ ـ ۴۳ـ ۳۲ ـ ۲۹ ـ ۲۷ـ ۲۲

☑ **پاسخ:** داده‌ها با تقریب ۱ گرد شده‌اند. $k = ۱ + ۳/۳۲۲\log_{۱۰}^{۲۵} = ۱ + ۳/۳۲(۱/۳۹) = ۵/۶۴ \simeq ۶$

توجه کنید که داده‌ها پیوسته‌اند. $s = \dfrac{۱}{۲} = ۰/۵$

$R = \max - \min = ۴۵/۵ - ۲۱/۵ = ۲۴$

$w = \dfrac{R}{k} = \dfrac{۲۴}{۶} = ۴$

توجه کنید که دامنه تغییرات از تفاضل کوچکترین داده واقعی از بزرگترین داده واقعی بدست می‌آید و چون دراینجا داده‌ها با تقریب ۱ گرد شده‌اند می‌توان گفت $\max = ۴۵/۵$ و $\min = ۲۱/۵$

حدود واقعی	x_i	f_i	r_i	g_i	s_i
۲۱/۵ − ۲۵/۵	۲۳/۵	۲	۰/۰۸	۲	۰/۰۸
۲۵/۵ − ۲۹/۵	۲۷/۵	۳	۰/۱۲	۵	۰/۲۰
۲۹/۵ − ۳۳/۵	۳۱/۵	۶	۰/۲۴	۱۱	۰/۴۴
۳۳/۵ − ۳۷/۵	۳۵/۵	۷	۰/۲۸	۱۸	۰/۷۲
۳۷/۵ − ۴۱/۵	۳۹/۵	۴	۰/۱۶	۲۲	۰/۸۸
۴۱/۵ − ۴۵/۵	۴۳/۵	۳	۰/۱۲	۲۵	۱

در اینجا x_i ها نماینده دسته (نشان‌دسته) نامیده می‌شوند که از جمع حد پایین با حد بالای هر طبقه تقسیم بر ۲ بدست می‌آید.

مثال ۷: *در یک جدول توزیع فراوانی با هشت طبقه و $n = ۱۰۰$ مجموع فراوانی نسبی تا قبل از طبقه هشتم برابر ۰/۹۲ است، فراوانی مطلق آخرین طبقه را بدست آورید.*

☑ **پاسخ:** $r_۸ = ۱ - ۰/۹۲ = ۰/۰۸ \Rightarrow r_۸ = \dfrac{f_۸}{n} \Rightarrow \dfrac{۸}{۱۰۰} = \dfrac{f_۸}{۱۰۰} \Rightarrow f_۸ = ۸$

تست ۲: اگر در یک جدول توزیع فراوانی حجم جامعه برابر ۴۰ و فراوانی مطلق طبقه سوم آن برابر ۵ باشد، درصد فراوانی نسبی آن چند درصداست؟

۱) ۰/۰۵ ۲) ۰/۱۲۵ ۳) ۸ ۴) ۱۲/۵

پاسخ: گزینه «۴»

$$r_i = \frac{f_i}{n} \Rightarrow r_۳ = \frac{f_۳}{n} = \frac{۵}{۴۰} = ۰/۱۲۵ \Rightarrow ۰/۱۲۵ \times ۱۰۰ = ۱۲/۵$$

تست ۳: اگر در یک جدول توزیع فراوانی با حجم نمونه ۲۵، فراوانی نسبی طبقه پنجم ۰/۴ باشد، فراوانی مطلق طبقه پنجم کدام است؟

۱) ۱۲ ۲) ۱۰ ۳) ۸ ۴) ۹

پاسخ: گزینه «۲»

$$r_۵ = \frac{f_۵}{n} \Rightarrow \frac{۴}{۱۰} = \frac{f_۵}{۲۵} \Rightarrow f_۵ = ۱۰$$

تست ۴: در یک جدول توزیع فراوانی، فراوانی تجمعی طبقه دوم و سوم برابر می‌باشد، کدام گزینه صحیح می‌باشد؟

۱) فراوانی مطلق طبقه دوم صفر است.

۲) فراوانی مطلق طبقه دوم و چهارم صفر است.

۳) فراوانی مطلق طبقه سوم صفر است.

۴) فراوانی مطلق طبقه دوم و چهارم برابر است.

پاسخ: گزینه «۳»

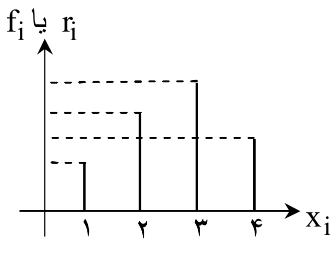
$$\begin{array}{l} g_۲ = f_۱ + f_۲ \\ g_۳ = f_۱ + f_۲ + f_۳ \end{array} \Rightarrow f_۳ = ۰$$

ترسیم نمودارهای گوناگون از روی مقادیر جدول

در اینجا به معرفی نمودارهایی برای داده‌های گسسته و پیوسته به صورت جداگانه می‌پردازیم.

نمودارهای میله‌ای و دایره‌ای برای داده‌های گسسته بکار می‌روند.

الف ـ نمودارهای میله‌ای: در این نمودار دو محور عمود بر هم را در نظر می‌گیریم و بر روی محور افقی مقادیر x_i ها و بر روی محور عمودی مقادیر فراوانی‌های مطلق (فراوانی‌های نسبی) را می‌نویسیم سپس روی هرمقدار x_i میله‌ای به ارتفاع فراوانی مطلق f_i (فراوانی نسبی r_i) رسم می‌کنیم.



تست ۵: در نمودار میله‌ای روی محور X ها چه اندازه‌هایی قرار می‌گیرند؟

۱) فراوانی نسبی ۲) فراوانی تجمعی ۳) فراوانی مطلق ۴) نشان دسته‌ها

پاسخ: گزینه «۴»

ب ـ نمودار دایره‌ای: در این نمودار دایره‌ای را رسم کرده و آن را به تعداد طبقات جدول فراوانی به قطعهایی تقسیم می‌کنیم. اندازه هر قطاع متناسب با فراوانی نسبی طبقه مربوطه می‌باشد.

درجه متناظر دسته i ام $= \frac{f_i}{n} \times ۳۶۰^\circ$

درصد متناظر دسته i ام $= \frac{f_i}{n} \times ۱۰۰$

در مثال ۶ برای طبقه دوم درجه قطاع مربوطه عبارت است از: $\frac{۳}{۲۵} \times ۳۶۰^\circ = ۴۳/۲^\circ$

تست ۶: در نمودار دایره‌ای ۱۰۰ داده آماری، کمانی به اندازه °۷۲ به یک طبقه تعلق دارد. فراوانی مطلق آن طبقه کدام است؟

۱) ۷ ۲) ۸ ۳) ۹ ۴) ۲۰

☑ **پاسخ: گزینه «۴»**

سهم برحسب درجه: $۷۲ = \frac{f_i}{۱۰۰} \times ۳۶۰ \Rightarrow ۳۶f_i = ۷۲۰ \Rightarrow f_i = ۲۰$

تست ۷: در نمودار دایره‌ای ۱۲۰ داده آماری، کمانی به اندازه °۶۰ به یک طبقه تعلق دارد فراوانی مطلق آن طبقه کدام است؟

۱) ۲۰ ۲) ۳۰ ۳) ۴۰ ۴) ۵۰

☑ **پاسخ: گزینه «۱»**

سهم بر حسب درجه $= \frac{f_i}{n} \times ۳۶۰ \Rightarrow ۶۰ = \frac{f_i}{۱۲۰} \times ۳۶۰ \Rightarrow f_i = ۲۰$

تست ۸: در نمودار دایره‌ای اگر فراوانی مطلق طبقه‌ای ۲۰ باشد و کمانی که برای این طبقه جدا شده برابر با ۲۰ درصد باشد مجموع فراوانی‌های مطلق کدام است؟

۱) ۱۰۰ ۲) ۱۱۰ ۳) ۱۵۰ ۴) ۷۵

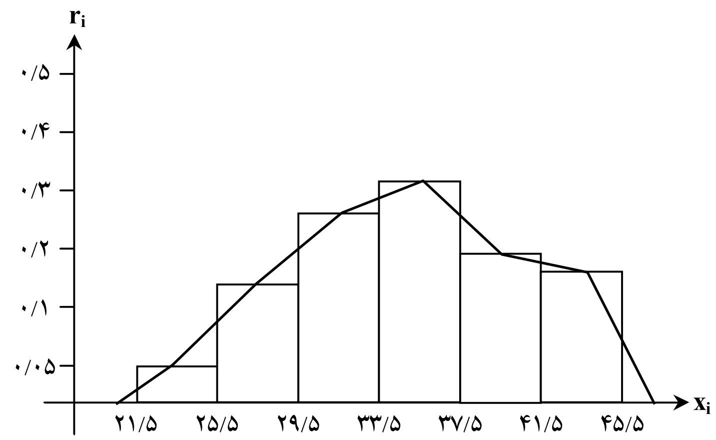
☑ **پاسخ: گزینه «۱»**

$\frac{۲۰}{۱۰۰} = \frac{۲۰}{n} \Rightarrow n = \sum f_i = ۱۰۰$

> **توجه :**
> ***نمودار دایره‌ای معمولاً برای داده‌ها کیفی نیز رسم می‌شود.***

نمودارهای هیستوگرام و چند بر فراوانی یا چند ضلعی برای داده‌های پیوسته بکار می‌روند.

الف ـ نمودار هیستوگرام: نموداری است متشکل از تعدادی مستطیل که تعداد این مستطیل‌ها برابر با تعداد رده‌های جدول فراوانی می‌باشد. قاعده هر مستطیل روی محور افقی قرار دارد و طول آن برابر با طول واقعی رده است. و مرکز آن نماینده رده است. ارتفاع مستطیل برابر فراوانی نسبی مربوط به آن رده یا فراوانی مطلق آن رده می‌باشد.



ب ـ چند بر فراوانی: اگر وسط قاعده‌های بالای مستطیل‌های هیستوگرام را بوسیله خطوط مستقیم به طور متوالی به یکدیگر وصل کرده و ابتدا و انتهای آن‌ها را به وسط ردهٔ ماقبل و ما بعد وصل کنیم یک چند ضلعی بوجود می‌آید که آن را چند بر فراوانی یا نمودار چندضلعی می‌نامند.

خلاصه کردن داده‌ها به یک یا چند عدد موسوم به شاخص یا آماره

با تشکیل جدول و رسم نمودار می‌توان تا حدودی اطلاعات در مورد داده‌ها به دست آورد. ولی برای آنکه بتوانیم نتایج کلی را به صورت ساده‌تر ارائه دهیم بهتر است که داده‌ها را در یک یا چند عدد خلاصه کنیم شاخصها خود به سه بخش شاخصهای تمرکز، شاخصهای پراکندگی و شاخصهای نسبی پراکندگی تقسیم می‌شوند.

شاخصهای تمرکز یا مرکزی:

این شاخص‌ها میزان تمرکز داده‌ها را نشان می‌دهند. شاخص‌های مهم مرکزی عبارتند از میانگین، میانه، مُد یا نما و چندک‌ها.

میانگین:

اگر داده‌ها بر روی یک محور به صورت منظم ردیف شوند، مقدار میانگین دقیقاً در نقطه تعادل یا مرکز ثقل توزیع قرار می‌گیرد. این محور همانند آلا کلنگ است که میانگین، نقطه تعادل آن است.

میانگین حسابی: یکی از مهمترین شاخص‌های تمرکز است. میانگین جامعه را با نماد μ و میانگین نمونه را با $\overline{x}$ نمایش می‌دهیم. فرض کنید داده‌های $x_1, x_2, ..., x_k$ به ترتیب دارای فراوانیهای $f_1, f_2, ..., f_k$ باشند و تعداد کل داده‌ها برابر با n باشد.

$$\overline{x} = \frac{1}{n}\sum_{i=1}^{k} f_i x_i$$

توجه شود اگر داده‌ها طبقه‌بندی شده باشند (داده‌های پیوسته) به جای x_i از نماینده طبقه استفاده می‌شود.

✍ **مثال ۸:** ***میانگین داده‌های زیررا بدست آورید.***

۲,۸,۴,۷,۶,۵,۹,۱,۳,۴

☑ **پاسخ :**

$$\overline{x} = \frac{1}{n}\sum_{i=1}^{k} x_i = \frac{1}{10}(2+8+4+7+6+5+9+1+3+4) = \frac{49}{10} = 4/9$$

مثال ۹: میانگین داده‌های روبرو را بدست آورید. ۲,۳,۳,۵,۵,۵,۶,۶,۲۰,۲۰,۲۰

پاسخ :

$$\bar{x}=\frac{1}{n}\sum_{i=1}^{k} f_i x_i=\frac{1}{11}(1\times 2+2\times 3+3\times 5+2\times 6+3\times 20)=8/64$$

مثال ۱۰: میانگین داده های جدول روبرو را بدست آورید.

x_i	۵	۱۰	۱۵	۲۰
f_i	۴	۸	۶	۷

پاسخ :

$$\bar{x}=\frac{\sum_{i=1}^{k} f_i x_i}{\sum_{i=1}^{k} f_i}=\frac{4\times 5+8\times 10+6\times 15+7\times 20}{25}=13/2$$

مثال ۱۱: میانگین داده‌های جدول زیر را بدست آورید.

حدود طبقات	۳-۵	۵-۷	۷-۹	۹-۱۱
f_i	۷	۴	۳	۵

پاسخ :

$$\bar{x}=\frac{1}{n}\sum_{i=1}^{4} f_i x_i=\frac{1}{7+4+3+5}(7\times 4+4\times 6+3\times 8+5\times 10)=6/631$$

تست ۹ : در جدول روبرو میانگین برابر با ۴/۹ است در اینصورت به جای نشان دسته سوم کدام عدد قرارمی‌گیرد؟

x_i	۱	۳	؟	۷	۹
f_i	۲	۴	۸	۵	۱

۱) ۵ ۲) ۴

۳) ۳ ۴) ۲

پاسخ: گزینه «۱»

$$\bar{x}=\frac{f_1x_1+f_2x_2+\cdots+f_nx_n}{f_1+f_2+\cdots+f_n}$$

$$4/9=\frac{1\times 2+3\times 4+8\times x+7\times 5+9\times 1}{2+4+8+5+1}\Rightarrow 98=8x+58\Rightarrow 8x=40\Rightarrow x=5$$

تست ۱۰: در جدول زیر میانگین کدام است؟

حدود طبقات	۲ − ۴	۵ − ۷	۸ − ۱۰
f_i	۲	۴	۲

۱) ۵ ۲) ۶

۳) ۷ ۴) ۸

پاسخ: گزینه «۲» ابتدا در جدول نشان دسته‌ها را بدست می‌آوریم:

x_i	۳	۶	۹
f_i	۲	۴	۲

$$\Rightarrow \bar{x}=\frac{f_1x_1+f_2x_2++f_3x_3}{f_1+f_2+f_3}=\frac{3\times 2+6\times 4+9\times 2}{2+4+2}=6$$

نکات مربوط به میانگین حسابی

۱ـ اگر داده‌های $x_1, x_2, \ldots, x_k$ تشکیل تصاعد عددی را بدهند میانگین آنها برابر است با: $\bar{x} = \dfrac{x_1 + x_k}{2}$

۲ـ مجموع انحرافات از میانگین برابر صفر است یعنی:

$$\sum_{i=1}^{n} (x_i - \bar{x}) = 0 \quad \text{یا} \quad \sum_{i=1}^{n} f_i (x_i - \bar{x}) = 0$$

۳ـ $\bar{x} = \dfrac{\sum_{i=1}^{n} x_i}{n} \Rightarrow \sum_{i=1}^{n} x_i = n\bar{x}$

۴ـ اگر به تمام داده‌ها مقدار ثابت a را اضافه یا کم کنیم به میانگین نیز مقدار a اضافه و یا کم می‌شود یعنی:

$$y_i = x_i \pm a \Rightarrow \bar{y} = \bar{x} \pm a$$

۵ـ هر گاه تمام داده‌ها را در عدد ثابتی مثل b ضرب یا تقسیم کنیم، میانگین نیز در b ضرب یا تقسیم می‌شود یعنی:

$$y_i = bx_i \Rightarrow \bar{y} = b\bar{x}$$

مثال ۱۲: ***اگر میانگین داده‌های $x_1, x_2, \ldots, x_n$ برابر با ۲۰ باشد میانگین داده‌های $6 + \frac{x_1}{3}, 6 + \frac{x_2}{3}, \ldots, 6 + \frac{x_n}{3}$ را بدست آورید.***

☑ **پاسخ:** $y_i = \dfrac{x_i}{3} + 6 \Rightarrow \bar{y} = \dfrac{\bar{x}}{3} + 6 = \dfrac{1}{3} \times 20 + 6 = 12/66$

۶ـ اگر $\bar{x}_1$ میانگین f_1 داده، $\bar{x}_2$ میانگین f_2 داده و $\cdots$، $\bar{x}_k$ میانگین f_k داده باشد میانگین جمیع این اعداد یعنی میانگین میانگین‌ها برابر است با:

$$\bar{x} = \frac{f_1\bar{x}_1 + f_2\bar{x}_2 + \ldots + f_k\bar{x}_k}{f_1 + f_2 + \ldots + f_k}$$

مثال ۱۳: ***دانشکده‌ای دارای ۵ رشته است که به ترتیب ۲۰ ، ۲۵ ، ۳۰ ، ۳۵ و ۳۲ نفر دانشجو دارند و میانگین درس عمومی ریاضی آنها به ترتیب ۱۸ ، ۱۷/۵ ، ۱۶/۸ ، ۱۶/۲ و ۱۷/۲ می‌باشد. میانگین درس ریاضی عمومی این دانشکده را بدست آورید.***

☑ **پاسخ:**

$$\bar{x} = \frac{20\times18 + 25\times17/5 + 30\times16/8 + 35\times16/2 + 32\times17/2}{20+25+30+35+32} = \frac{360 + 437/5 + 504 + 567 + 550/4}{142} = 17/034$$

تست ۱۱: ***میانگین ۴ عدد ۸۰ است اگر به این چهار عدد، عدد ۳۰ اضافه شود، میانگین ۵ عدد حاصل کدام است؟***

۱) ۵۵ ۲) ۶۰ ۳) ۷۰ ۴) ۷۵

☑ **پاسخ: گزینه «۳»**

$$\frac{x_1 + x_2 + x_3 + x_4}{4} = 80 \Rightarrow x_1 + x_2 + x_3 + x_4 = 320$$

$$\frac{x_1 + x_2 + x_3 + x_4 + 30}{5} = \frac{320 + 30}{5} = 70$$

تست ۱۲: ***اگر میانگین داده‌های $x_1, x_2, \dots, x_n$ برابر با ۱۰ باشد و هر یک از داده‌ها را در عدد ۳ ضرب و با عدد ۵ جمع کنیم، میانگین داده‌های جدید کدام است؟***

۱) ۱۰ ۲) ۱۵ ۳) ۳۰ ۴) ۳۵

پاسخ: گزینه «۴»

$$\frac{x_1 + x_2 + x_3 + \cdots + x_n}{n} = 10$$

عملی که بر روی داده‌ها انجام می‌شود بر روی میانگین نیز انجام می‌شود.

$$\bar{y} = 3 \times 10 + 5 = 35$$

تست ۱۳: ***در صورتیکه میانگین $x_1, x_2, \dots, x_{30}$ برابر با a باشد در این صورت حاصل $\sum_{i=1}^{30}(x_i - a)$ کدام است؟***

۱) ۳۰ ۲) $30a$ ۳) $15a$ ۴) صفر

پاسخ: گزینه «۴» مجموع انحراف داده‌ها از میانگین همواره صفر است.

میانگین وزنی یا موزون

فرض کنید شما می‌خواهید معدل دروس خود را در یک ترم محاسبه کنید. برای درک بهتر به مثال زیر توجه کنید.

مثال ۱۴: ***دانشجویی در ترم گذشته در درس ریاضی نمره ۱۵، در درس آمار نمره ۱۲ و در درس ادبیات ۱۴ گرفته است اگر ریاضی و آمار هر کدام ۴ واحد و ادبیات درس ۳ واحدی باشد معدل او در ترم گذشته چقدر بوده است؟***

پاسخ:

$$\bar{x} = \frac{4 \times 15 + 4 \times 12 + 3 \times 14}{4 + 4 + 3} = \frac{150}{11} = 13/64$$

پس در حالت کلی اگر داده‌های $x_1, x_2, \dots, x_n$ به ترتیب دارای ضرایب وزنی $w_1, w_2, \dots, w_n$ باشد آنگاه:

$$\text{میانگین وزنی} = \frac{\sum_{i=1}^{n} w_i x_i}{\sum_{i=1}^{n} w_i}$$

پس هنگامی که هر یک از داده‌ها، وزنی متناسب داشته باشند برای محاسبه میانگین به هر داده وزن خودش داده می‌شود.

میانه

در داده‌های آماری عددی که از پنجاه درصد داده‌ها کوچکتر و از پنجاه درصد داده‌ها بزرگتر باشد را **میانه** گویند. میانه را با نماد Md نمایش می‌دهند.

محاسبه میانه برای داده‌های گسسته

ابتدا داده‌ها را به صورت صعودی مرتب می‌کنیم اگر تعداد داده‌ها فرد باشد میانه عدد وسط می‌باشد و اگر تعداد داده‌ها زوج باشد میانه برابر با میانگین دو عدد وسط می‌باشد.

n فرد باشد $\quad Md = x_{\frac{n+1}{2}}$

n زوج باشد $\quad Md = \dfrac{x_{\frac{n}{2}} + x_{\frac{n}{2}+1}}{2}$

مثال ۱۵: میانه را برای داده‌های ۱۲ و۴ و ۸ و ۷ و ۵ بدست آورید.

☑ **پاسخ:** ابتدا داده‌ها را به صورت صعودی مرتب می‌کنیم. $(n=5)$

$$4,5,7,8,12 \Rightarrow Md = 7$$

مثال ۱۶: میانه را برای داده‌های ۸ و ۱۱ و۳ و۹ و ۵ و۴ بدست آورید.

☑ **پاسخ:** ابتدا داده‌ها را به صورت صعودی مرتب می‌کنیم. $(n=6)$

$$3,4,5,8,9,11 \Rightarrow Md = \frac{5+8}{2} = 6/5$$

اکنون اگر داده‌های ما به صورت گسسته همراه با فراونی ارائه شوند برای محاسبه میانه ابتدا ستون فراوانی تجمعی را تشکیل می‌دهیم اگر تعداد داده‌ها فرد باشد عدد $k=\frac{n+1}{2}$ و اگر تعداد داده‌ها زوج باشد $k=\frac{n}{2}$ می‌گیریم و مقدار k بین دو فراوانی تجمعی یا مساوی یکی از فراوانی‌های تجمعی است در اینصورت:

$$g(x_{i-1}) < k \le g(x_i) \Rightarrow Md = x_i \qquad n \text{ فرد}$$

$$\left.\begin{array}{l} g(x_{i-1}) < k < g(x_i) \Rightarrow Md = x_i \\ k = g(x_i) \Rightarrow Md = \frac{1}{2}(x_i + x_{i+1}) \end{array}\right\} n \text{ زوج}$$

برای درک بهتر به مثالهای زیر توجه کنید.

مثال ۱۷: در جدول روبرو، میانه را بدست آورید.

x_i	f_i
۱۲	۴
۱۴	۸
۱۵	۶
۱۷	۴
۱۹	۳

☑ **پاسخ:** ابتدا باید فراوانی تجمعی بدست آید.

x_i	۱۲	۱۴	۱۵	۱۷	۱۹
f_i	۴	۸	۶	۴	۳
g_i	۴	۱۲	۱۸	۲۲	۲۵

$$k = \frac{n+1}{2} = \frac{25+1}{2} = 13 \Rightarrow 12 = g(x_2) < k = 13 < g(x_3) = 18 \Rightarrow Md = x_3 = 15$$

مثال ۱۸: در جدول زیر میانه را بدست آورید.

x_i	۱۴	۱۵	۱۷	۱۸	۱۹
f_i	۳	۶	۵	۲	۴

☑ **پاسخ:** ابتدا فراوانی تجمعی را بدست می‌آوریم.

x_i	۱۴	۱۵	۱۷	۱۸	۱۹
f_i	۳	۶	۵	۲	۴
g_i	۳	۹	۱۴	۱۶	۲۰

$$k = \frac{n}{2} = \frac{20}{2} = 10$$

$$g(x_2) = 9 < k = 10 < 14 = g(x_3) \Rightarrow Md = x_3 = 17$$

مثال ۱۹: در جدول زیر میانه را بدست آورید.

x_i	۱۴	۱۵	۱۷	۱۸	۱۹
f_i	۴	۶	۵	۲	۳
g_i	۴	۱۰	۱۵	۱۷	۲۰

پاسخ:

$$k = \frac{n}{۲} = \frac{۲۰}{۲} = ۱۰$$

$$k = ۱۰ = g(x_۲) \Rightarrow Md = \frac{x_۲ + x_۳}{۲} = \frac{۱۵+۱۷}{۲} = ۱۶$$

محاسبه میانه برای داده‌های پیوسته

برای محاسبه میانه ابتدا ستون فراوانی تجمعی را تشکیل می‌دهیم. اولین طبقه‌ای که فراوانی تجمعی آن بزرگتر یا مساوی $\frac{n}{۲}$ باشد را به عنوان طبقه میانه‌دار انتخاب و میانه را از فرمول زیر محاسبه می‌کنیم.

L: کران پایین واقعی f_i: فراوانی مطلق طبقه میانه دار

$g_{i-۱}$: فراوانی تجمعی طبقه قبل w : طول دسته

$$Md = L + \left(\frac{\frac{n}{۲} - g_{i-۱}}{f_i}\right) \times w$$

مثال ۲۰: در جدول زیر مقدار میانه را بدست آورید.

حدود طبقات	۵-۸	۹-۱۲	۱۳-۱۶	۱۷-۲۰
حدود واقعی طبقات	۴/۵-۸/۵	۸/۵-۱۲/۵	۱۲/۵-۱۶/۵	۱۶/۵-۲۰/۵
f_i	۵	۱۲	۱۴	۱۰
g_i	۵	۱۷	۳۱	۴۱

پاسخ: $\frac{n}{۲} = ۲۰/۵$ پس طبقه سوم طبقه میانه‌دار است. چون فراوانی تجمعی آن بزرگتر یا مساوی ۲۰/۵ است.

$$L = ۱۲/۵ \;;\; g_{i-۱} = ۱۷ \;;\; f_i = ۱۴ \;;\; w = ۴$$

$$Md = ۱۲/۵ + \frac{۲۰/۵ - ۱۷}{۱۴} \times ۴ = ۱۳/۵$$

نکاتی مربوط به میانه

۱ـ در هر جامعه آماری فقط یک میانه وجود دارد.

۲ـ $\sum |x_i - Md|$ همواره مینیمم می‌باشد.

۳ـ بر خلاف میانگین، میانه از اعداد خیلی بزرگ و خیلی کوچک داده‌ها متأثر نمی‌شود.

۴ـ از نظر هندسی میانه خطی عمود به معادله $x = Md$ است که نمودار هیستوگرام را از نظر سطح به دو قسمت مساوی تقسیم می‌کند.

تست ۱۴: مقدار $\sum_{i=1}^{n} |x_i - A|$ زمانی مینیمم است که:

۱) $A = Md$ ۲) $A = \overline{x}$ ۳) $A = Mo$ ۴) $A = Q_۱$

پاسخ: گزینه «۱» طبق نکته گفته شده در متن درس این مقدار زمانی کمترین مقدار است که $A = Md$

مُد یا نما: اندازه‌ای از صفت می‌باشد که فراوانی آن بیشترین مقدار را دارا باشد و آن را با نماد Mo نمایش می‌دهند.

توجه :
اگر فراوانی داده‌ها برابر باشند داده‌ها مُد ندارند.

توجه :
در یک جامعه آماری ممکن است بیش از یک مُد داشته باشیم.

مثال ۲۱: ***در داده‌های روبرو مُد را بیابید.*** ***۴ و۲ و۳ و ۴ و۷ و۶ و۵ و۴***

☑ **پاسخ:** بیشترین فراوانی متعلق به عدد ۴ است. پس $Mo = ۴$.

مثال ۲۲: ***در داده‌های روبرو مُد را بیابید.*** ***۵ و۹ و۵ و۴ و۶ و۹ و ۶ و۹ و۷ و۵***

☑ **پاسخ:** دو عدد ۵ و ۹ بیش از همه تکرار شده‌اند پس ۹ و۵ = Mo در نتیجه داده‌ها دو نمایی هستند.

مثال ۲۳: ***در داده‌های روبرو مُد را بیابید.*** ***۶ و۴ و۵ و۶ و۵ و۴ و۴ و۵ و۶***

☑ **پاسخ:** فراوانی هر سه داده یکسان است پس داده‌ها مُد ندارند.

تست ۱۵: ***اگر مشاهدات ۸ ,۸ ,۸ ,۸ ,۸ باشند مقدار مُد کدام است؟***

۱) ۸ ۲) ۱ ۳) صفر ۴) داده‌ها مُد ندارند.

☑ **پاسخ: گزینه** «۴» برای داده‌های مساوی مُد وجود ندارد.

محاسبه مُد در داده‌های پیوسته: ابتدا طبقه‌ای که فراوانی آن بیشتر است را مشخص می‌کنیم سپس از فرمول زیر مُد را بدست می‌آوریم.

$$Mo = L + \left(\frac{d_۱}{d_۱ + d_۲}\right).w$$

L: کران پائین واقعی طبقه مُد دار

$d_۱ : f_i - f_{i-۱}$: تفاضل فراوانی مطلق طبقه قبل از طبقه مُددار

$d_۲ : f_i - f_{i+۱}$: تفاضل فراوانی مطلق طبقه بعد از طبقه مُددار

W: طول طبقه

مثال ۲۴: ***در جدول زیر مُد را بدست آورید.***

حدود طبقات	۲-۵	۶-۹	۱۰-۱۳	۱۴-۱۷
f_i	۷	۶	۱۰	۸

☑ **پاسخ:** طبقه سوم بیشترین فراوانی را دارد پس در طبقه سوم مُد وجود دارد. $w = ۴$; $d_۲ = ۲$; $d_۱ = ۴$; $L = ۹/۵$

$$Mo = ۹/۵ + \frac{۴}{۴+۲} \times ۴ = ۱۲/۱۶۷$$

تست ۱۶: در جدول زیر مُد کدام است؟

حدود واقعی طبقات	۲/۵-۵/۵	۵/۵-۸/۵	۸/۵-۱۱/۵
f_i	۴	۲۰	۱۲

۱) ۶/۵ ۲) ۷/۵ ۳) ۸/۵ ۴) ۹/۵

☑ **پاسخ: گزینه «۲»** طبق فرمول زیر برای داده‌های پیوسته مُد بدست می‌آید. ابتدا طبقه‌ای که بیشترین فراوانی را دارد مشخص می‌کنیم(طبقه دوم)

$$MO = L + \left(\frac{d_۱}{d_۱ + d_۲}\right).w = ۵/۵ + \left(\frac{۱۶}{۱۶+۸}\right) \times ۳ = ۷/۵$$

تست ۱۷: در جدول زیر مد کدام است؟

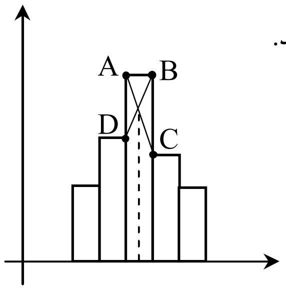
حدود طبقات	۱۰۰-۱۰۹	۱۱۰-۱۱۹	۱۲۰-۱۲۹	۱۳۰-۱۳۹
f_i	۵	۶	۱۰	۴

۱) ۱۲۵/۵ ۲) ۱۲۲/۵ ۳) ۱۲۳/۵ ۴) ۱۲۴/۵

☑ **پاسخ: گزینه «۳»** ابتدا طبقه ای که فراونی آن بیشترین است را انتخاب می‌کنیم (طبقه سوم)

$$MO = L + \left(\frac{d_۱}{d_۱ + d_۲}\right) \times w = ۱۱۹/۵ + \left(\frac{۴}{۴+۶}\right) \times ۱۰ = ۱۲۳/۵$$

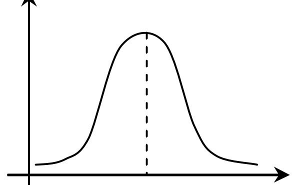
محاسبه مُد به روش ترسیمی: ابتدا هیستوگرام را رسم می‌کنیم بلندترین مستطیل رامطابق شکل درنظر گرفته و ازA به C و از B به D وصل می‌کنیم. از محل تلاقی این دو خط بر محور xها عمود می‌کنیم پای عمود همان مُد یا نما می‌باشد.



توجه :
در بین شاخص‌های مرکزی مُد اهمیت زیادی ندارد ولی در داده‌های کیفی، مُد تنها شاخص مرکزی است.

رابطه تجربی بین میانگین ، میانه و مُد

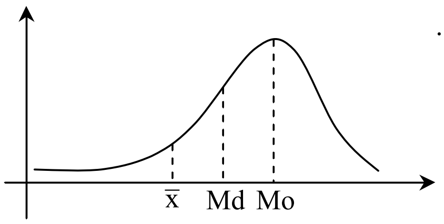
یک توزیع را متقارن گویند اگر منحنی فراوانی آن به فرم زنگی شکل باشد. در توزیع‌های متقارن، میانگین و میانه و مُد بر هم منطبق هستند.



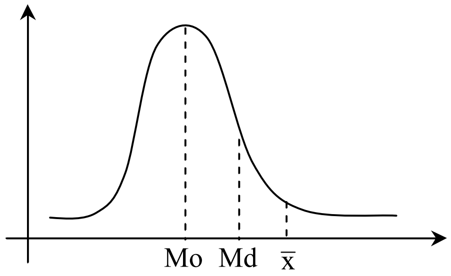
$\bar{x} = Md = Mo$

توجه کنید که اگر در یک نمونه مقادیر میانگین، میانه و مُد مساوی باشند دلیل بر تقارن توزیع داده‌ها نیست. توزیع‌هایی که کمی از تقارن خارج شده‌اند را توزیع‌های چوله می‌نامند. چولگی ممکن است به چپ یا به راست باشد.

در توزیع‌های چوله به چپ $\bar{x} < Md < Mo$



در توزیع‌های چوله به راست $Mo < Md < \overline{x}$



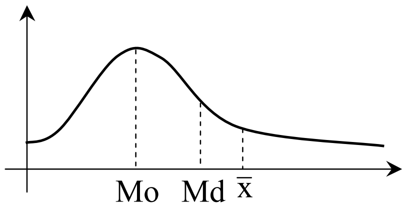
در توزیع‌هایی که چولگی زیاد نباشد، رابطه‌ای تجربی به نام رابطه پیرسن برقرار است.

$$\overline{x} - Mo \approx ۳(\overline{x} - Md)$$

تست ۱۸: ***در صورتی که جامعه‌ای دارای چولگی به راست باشد کدام رابطه صحیح است؟***

۱) $\overline{x} > Md > Mo$ ۲) $\overline{x} < Md < Mo$ ۳) $\overline{x} = Md = Mo$ ۴) هیچکدام

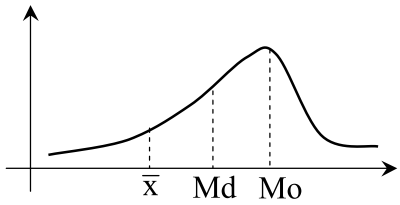
☑ **پاسخ: گزینه «۱»** $\overline{x} > Md > Mo$



تست ۱۹: ***در منحنی‌هایی که دارای چولگی به سمت چپ می‌باشند، کدام گزینه صحیح است؟***

۱) $MO > \overline{x} > Md$ ۲) $MO < x < Md$ ۳) $MO > Md > \overline{x}$ ۴) $MO < Md < \overline{x}$

☑ **پاسخ: گزینه «۳»** $Mo > Md > \overline{x}$



تست ۲۰: ***میانگین و میانه در جامعه‌ای به ترتیب ۳۰ و ۲۵ می‌باشد در اینصورت اگر توزیع جامعه از چولگی معقولی برخوردار باشد مُد کدام است؟***

۱) ۲۵ ۲) ۳۰ ۳) ۱۵ ۴) ۹۰

☑ **پاسخ: گزینه «۳»** از رابطه پیرسون استفاده می‌کنیم:

$$\overline{x} - MO = ۳(\overline{x} - Md)$$

$$۳۰ - MO = ۳(۳۰ - ۲۵) \Rightarrow MO = ۱۵$$

چندک‌ها

مقادیری می‌باشند که دامنه تغییرات داده‌هایی را که به صورت صعودی مرتب شده‌اند به فواصل چندکی مورد نیاز تقسیم می‌کنند به طوریکه فراوانی‌ها در هر یک از این فواصل، درصد مشخصی از کل فراوانی را تشکیل می‌دهند. چندکهای معروف عبارتند از:

الف ـ چارکها: که دامنه تغییرات را به چهار قسمت مساوی تقسیم می‌کنند و آنها را با $Q_۱, Q_۲, Q_۳$ نشان می‌دهیم. توجه کنید که $Q_۲$ همان میانه است.

ب ـ دهک‌ها: که دامنه تغییرات را به ده قسمت مساوی تقسیم می‌کنند و آنها را با $D_۱, D_۲, D_۳, ..., D_۹$ نمایش می‌دهیم. توجه کنید $D_۵$ همان میانه است.

ج ـ صدک‌ها: که دامنه تغییرات را به صد قسمت مساوی تقسیم می‌کنند و آنها را با $P_۱, P_۲, ..., P_{۹۹}$ نشان می‌دهیم. توجه کنید که $P_{۵۰}$ همان میانه است.

ما می‌توانیم از روی صدکها، چارکها و دهکها را بدست آوریم. مثلاً صدک ۷۵ام همان چارک سوم یا دهک چهارم همان صدک ۴۰ ام است.

محاسبه چندکها در داده‌های گسسته

فرض کنید n داده را به صورت صعودی مرتب کرده‌ایم برای محاسبه صدک ۱۰۰P ام، ابتدا $(n+۱)P$ را محاسبه می‌کنیم اگر این مقدار عدد صحیحی مانند k شود در اینصورت x_k همان صدک ۱۰۰P ام می‌باشد ولی اگر $(n+۱)P$ عدد صحیحی نشد مقدار صحیح آن را k و اعشار آن را که r است جدا می‌کنیم و صدک ۱۰۰P ام را از فرمول زیر محاسبه می‌کنیم.

$$\text{صدک ۱۰۰P ام} = (۱-r)x_k + rx_{k+۱}$$

برای درک بهتر به مثالهای زیر توجه کنید.

مثال۲۵: در داده‌های ۱۹, ۱۳, ۱۴, ۱۸, ۱۶, ۱۵, ۱۷ مقدار چارک اول و چارک دوم را بدست آورید.

☑ **پاسخ:** ابتدا داده‌ها را به صورت صعودی مرتب می‌کنیم و مقدار $(n+۱)P$ را محاسبه می‌کنیم.

۱۳,۱۴,۱۵,۱۶,۱۷,۱۸,۱۹

چارک اول همان صدک ۲۵ ام است پس: $P = ۰/۲۵$

چارک دوم همان صدک ۵۰ ام است پس: $P = ۰/۵۰$

$$(n+۱).P = (۷+۱)\times ۰/۲۵ = ۲$$

$\Rightarrow x_۲ = ۱۴$ چارک اول است

$(n+۱).P = (۷+۱)\times ۰/۵ = ۴ \Rightarrow x_۴ = ۱۶$ چارک دوم است

تست ۲۱: در داده‌های روبرو دهک هفتم کدام است؟ ***۳, ۴, ۴, ۵, ۷, ۸, ۹, ۹, ۱۰***

۱) ۸/۵ ۲) ۷ ۳) ۹ ۴) ۸

☑ **پاسخ: گزینه «۳»** یعنی داده هفتم

$$(n+۱)P = (۹+۱)\times \frac{۷}{۱۰} = ۷ \Rightarrow D_۷ = x_۷ = ۹$$

مثال ۲۶: در داده‌های روبرو دهک هفتم را بدست آورید. ***۱۲, ۱۴, ۱۷, ۱۸, ۱۹, ۲۱***

☑ **پاسخ:** دهک هفتم یعنی صدک هفتادم پس: $P = ۰/۷$

$$(n+۱).P = ۷\times ۰/۷ = ۴/۹ \Rightarrow k = ۴$$

$$r = ۰/۹$$

$$\text{صدک هفتادم} = (۱-۰/۹)x_۴ + ۰/۹x_۵ = ۰/۱\times ۱۸ + ۰/۹\times ۱۹ = ۱/۸ + ۱۷/۱ = ۱۸/۹$$

محاسبه چندکها در داده‌های پیوسته

برای محاسبه صدک ۱۰۰P ام، عدد np را با ستون فراوانی تجمعی مقایسه می‌کنیم اولین طبقه‌ای که فراوانی تجمعی آن بزرگتر یا مساوی np باشد طبقه صدک ۱۰۰P ام می‌باشد و این صدک از فرمول زیر محاسبه می‌شود.

$$\text{صدک ۱۰۰P ام} = L + \left(\frac{np - g_{i-۱}}{f_i}\right).w$$

L : کران پائین واقعی مربوط به طبقه صدک

$g_{i-۱}$: فراوانی تجمعی طبقه ماقبل

f_i : فراوانی مطلق مربوط به طبقه صدک

w : طول طبقه

مثال ۲۷: در جدول زیر صدک ۶۸ام، دهک هفتم، چارک سوم را بدست آورید.

حدود طبقات	۴-۶	۷-۹	۱۰-۱۲	۱۳-۱۵
f_i	۸	۷	۱۰	۱۲
g_i	۸	۱۵	۲۵	۳۷

☑ **پاسخ:** صدک ۶۸ام یعنی $P = ۰/۶۸$ پس $n \times P = ۳۷ \times ۰/۶۸ = ۲۵/۱۶$ پس طبقه آخر طبقه صدک ۶۸ام است.

$$\text{صدک ۶۸ام} = ۱۲/۵ + \frac{۲۵/۱۶ - ۲۵}{۱۲} \times ۳ = ۱۲/۵۴$$

دهک هفتم: $n \times P = ۳۷ \times ۰/۷ = ۲۵/۹$، طبقه آخر

$$\text{دهک هفتم} = ۱۲/۵ + \frac{۲۵/۹ - ۲۵}{۱۲} \times ۳ = ۱۲/۷۲۵$$

چارک سوم: $n \times P = ۳۷ \times ۰/۷۵ = ۲۷/۷۵$

$$\text{چارک سوم} = ۱۲/۵ + \frac{۲۷/۷۵ - ۲۵}{۱۲} \times ۳ = ۱۳/۱۹$$

تست ۲۲: چارک اول داده‌های جدول روبرو کدام است؟

حدود طبقات	۱-۵	۵-۹	۹-۱۳	۱۳-۱۷
f_i	۹	۸	۱۱	۱۴

۱) $\frac{۴۷}{۸}$ ۲) $\frac{۴۳}{۴}$

۳) $\frac{۴۲}{۷}$ ۴) $\frac{۲۳}{۴}$

☑ **پاسخ: گزینه «۴»**

$$n \times p = ۴۲ \times \frac{۲۵}{۱۰۰} = ۱۰/۵$$

اولین طبقه‌ای که فراوانی تجمعی آن بزرگتر یا مساوی ۱۰/۵ است طبقه دوم است. پس:

$$\text{چارک اول} = L + \left(\frac{np - g_{i-۱}}{f_i}\right) \times w = ۵ + \left(\frac{۱۰/۵ - ۹}{۸}\right) \times ۴ = \frac{۲۳}{۴}$$

شاخصهای پراکندگی

پس از بحث و بررسی بر روی میزان تمرکز داده‌ها در اینجا به بیان معیارهایی برای تعیین میزان پراکندگی داده‌ها می‌پردازیم.

دامنه تغییرات

عبارت است از تفاضل کوچکترین داده از بزرگترین داده.

$$\begin{cases} R = \max - \min & \text{در داده‌های گسسته} \\ R = \max - \min + \alpha & \text{در داده‌های پیوسته} \end{cases}$$

که α تقریب است.

مثال ۲۸: در داده‌های زیر دامنه هر گروه را بدست آورید. (داده‌ها گسسته‌اند)

گروه (۱) ۲, ۱۷, ۱۷, ۱۷, ۱۷, ۱۷, ۱۸ $R_۱ = ۱۸ - ۲ = ۱۶$

گروه (۲) ۳, ۴, ۵, ۶, ۱۲, ۱۴, ۱۵ $R_۲ = ۱۵ - ۳ = ۱۲$

توجه شود که در گروه اول دامنه بزرگتر است ولی گروه دوم پراکندگی بیشتری دارد. پس دامنه جهت مقایسه پراکندگی داده‌ها معیار مناسبی نیست.

انحراف چارکها (نیم دامنه چارکها)

انحراف چارکها به صورت زیر تعریف می‌شود:

$$Q = \frac{Q_3 - Q_1}{2}$$

مثال ۲۹: ***برای داده‌های روبرو، نیم دامنه چارکها را محاسبه کنید.*** ***۳, ۴, ۷, ۸, ۹, ۱۰, ۱۲, ۱۴, ۱۶***

پاسخ:

$$(n+1)P = (9+1)\times 0/25 = 2/5 \Rightarrow k = 2\ ,\ r = 0/5$$

$$Q_1 = (1-r)x_k + rx_{k+1} = (1-0/5)\times 4 + 0/5\times 7 = 5/5$$

$$(n+1)p = (9+1)\times 0/75 = 7/5 \Rightarrow k = 7, r = 0/5$$

$$Q_3 = (1-0/5)\times 12 + 0/5\times 14 = 13$$

$$Q = \frac{1}{2}(Q_3 - Q_1) = \frac{1}{2}(13 - 5/5) = 3/75$$

تست ۲۳: ***اگر $Md = 25$, $Q_1 = 20$, $Q_3 = 30$ باشد در این حالت انحراف چارکی کدام است؟***

۱) ۲۰ ۲) ۵ ۳) ۱۰ ۴) ۴

پاسخ: گزینه «۲»

$$Q = \frac{Q_3 - Q_1}{2} = \frac{30 - 20}{2} = 5$$

انحراف از میانگین

معیار دیگر پراکندگی، انحراف از میانگین است دامنه تغییرات و انحراف چارکها این عیب را دارند که به تمام داده‌ها بستگی ندارند.

$$A.D = \frac{1}{n}\sum_{i=1}^{k} f_i\left|x_i - \overline{x}\right|$$

نکاتی در مورد انحراف از میانگین

۱ـ اگر انحراف از میانگین صفر باشد، آنگاه همه متغیرها با هم برابرند و برعکس یعنی:

$$A.D = 0 \Leftrightarrow x_1 = x_2 = \ldots = x_n$$

۲ـ اگر به تمام داده‌ها عدد ثابت a را اضافه یا کم کنیم، انحراف از میانگین آنها تغییری نمی‌کند.

$$y_i = x_i \pm a \Rightarrow A.D_y = A.D_x$$

۳ـ اگر تمام داده‌ها را در عدد ثابت a ضرب کنیم انحراف از میانگین در قدر مطلق a ضرب می‌شود یعنی:

$$y_i = ax_i \Rightarrow AD_y = |a|.A.D_x$$

مثال ۳۰: ***اگر انحراف از میانگین اعداد a, b, c, d و ۵ برابر صفر باشد میانگین اعداد $2a, 4b, 3c, 2d$ را بدست آورید.***

پاسخ: طبق نکته ۱:

$$A.D = 0 \Rightarrow a = b = c = d = 5$$

$$\overline{x} = \frac{2\times5 + 4\times5 + 3\times5 + 2\times5}{4} = \frac{55}{4} = 13/75$$

واریانس و انحراف معیار (انحراف استاندارد)

مهمترین شاخص برای انحراف از میانگین واریانس است. واریانس نمونه را با $s^۲$ و واریانس جامعه را با $\sigma^۲$ نمایش می‌دهند.

$$Var(x)=\frac{۱}{n}\sum_{i=۱}^{n} f_i(x_i-\overline{x})^۲=\sum_{i=۱}^{n}\frac{f_i x_i^۲}{n}-(\overline{x})^۲$$

به جذر مثبت واریانس انحراف معیار می‌گویند.

مثال ۳۱: ***واریانس اعداد ۶, ۵, ۴, ۳, ۲ را محاسبه کنید.***

☑ پاسخ:

$$\overline{x}=\frac{\sum_{i=۱}^{n} f_i x_i}{n}=\frac{۲+۳+۴+۵+۶}{۵}=۴$$

$$Var(x)=s^۲=\frac{\sum_{i=۱}^{n} f_i x_i^۲}{n}-(\overline{x})^۲=\frac{(۲)^۲+(۳)^۲+(۴)^۲+(۵)^۲+(۶)^۲}{۵}-(۴)^۲=۲$$

> **توجه :**
> ***عیب واریانس آن است که نمی‌توان آنرا مقایسه کرد.***

نکات واریانس و انحراف معیار

۱ـ اگر تمام داده‌ها را در عدد ثابتی مانند a ضرب یا تقسیم کنیم و به آنها عدد b را اضافه کنیم در اینصورت واریانس جدید برابر با مربع ضریب در واریانس قبلی است.

$$var(ax\pm b)=a^۲\ var(x)$$

۲ـ اگر تمام داده‌ها را در عدد ثابتی مانند a ضرب یا تقسیم کنیم در اینصورت انحراف معیار در قدر مطلق a ضرب می‌شود.

$$\sigma_{ax}=|a|\sigma_x$$

۳ـ اگر تمام داده‌ها مساوی باشند واریانس و انحراف معیار صفر است و برعکس یعنی:

$$x_۱=x_۲=...=x_n \Leftrightarrow var(x)=۰$$

$$x_۱=x_۲=...=x_n \Leftrightarrow \sigma_x=۰$$

۴ـ $var(x)=\overline{x^۲}-\overline{x}^۲$

تست ۲۴: ***اگر واریانس اعداد $x_۱, x_۲, ..., x_k$ برابر با a باشد واریانس $۲x_۱+۱, ۲x_۲+۱, ..., ۲x_k+۱$ کدام است؟***

۱) $۴a^۲$ ۲) $\frac{۴}{a}$ ۳) $۴a$ ۴) $۴a^۳$

☑ پاسخ: گزینه «۳»

$$Var(x_۱,x_۲,\ldots,x_k)=a$$

$$Var(۲x_۱+۱,۲x_۲+۱,\ldots,۲x_k+۱)=۴\ var(x_۱,x_۲,\ldots,x_k)=۴a$$

چون

$$Var(ax+b)=a^۲\ var(x)$$

***تست ۲۵:** اگر $Var(x)$ برابر با ۴ باشد $Var(ax+b)$ کدام است؟*

۱) $2a$ ۲) $3a$ ۳) $4a^2$ ۴) $8\sqrt{a}$

☑ **پاسخ: گزینه «۳»** $Var(ax+b)=a^2Var(x)=4a^2$

***تست ۲۶:** اگر s انحراف معیار داده‌های $x_1, x_2, \ldots, x_{20}$ باشد، انحراف معیار داده‌های $2x_1+5, 2x_2+5, \ldots, 2x_{20}+5$ کدام است؟*

۱) $2s$ ۲) $2s+5$ ۳) $2s+10$ ۴) $2s+100$

☑ **پاسخ: گزینه «۱»** $\sigma(2x_1+5, 2x_2+5, \cdots 2x_{20}+5)=2\sigma(x_1, x_2, \cdots x_{20})=2s$

اندازه استاندارد

اگر $x_1, x_2, \ldots, x_n$ مقادیر مشاهده شده باشند و از هر داده میانگین $(\bar{x})$ را کم کرده و حاصل را بر انحراف معیار s تقسیم کنیم، متغیر جدیدی بوجود می‌آید که آن را با z_i نمایش می‌دهند و به آن **متغیر استاندارد** گویند.

$$z_i=\frac{x_i-\bar{x}}{s}$$

***مثال ۳۲:** فرض کنید میانگین نمرات شما و دوستانتان در درس ادبیات ۵۷ و انحراف معیار ۷ باشد اگر نمره شما در درس ادبیات ۷۲ باشد نمره استاندارد شما در درس ادبیات چقدر است؟*

☑ **پاسخ:** نمره استاندارد شده در درس ادبیات $z=\frac{72-57}{7}=2/14$

گشتاورها

فرض کنید $x_1, x_2, \ldots, x_n$ مقادیر مشاهده متغیر x باشد گشتاور مرکزی مرتبه rاُم به صورت زیر تعریف می‌شود:

$$m_r=\frac{1}{n}\sum_{i=1}^{n} f_i(x_i-\bar{x})^r$$

***مثال ۳۳:** برای داده‌های ۲، ۳، ۷، ۸، ۱۰ گشتاورهای مرتبه اول و دوم را بدست آورید.*

☑ **پاسخ:**

$$\bar{x}=\frac{1}{5}\sum_{i=1}^{5}x_i=\frac{1}{5}(2+3+7+8+10)=6$$

$$m_1=\frac{1}{5}\sum_{i=1}^{n}(x_i-\bar{x})=\frac{1}{5}\left[(2-6)+(3-6)+(7-6)+(8-6)+(10-6)\right]=0$$

$$m_2=\frac{1}{5}\sum_{i=1}^{n}(x_i-\bar{x})^2=\frac{1}{5}\left[(2-6)^2+(3-6)^2+(7-6)^2+(8-6)^2+(10-6)^2\right]=\frac{46}{5}=9/2$$